

PUBLICATIONS DE
L'INSTITUT de
STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

CONTRIBUTION A LA THEORIE DE
L'ESTIMATION FONCTIONNELLE

D. BOSQ

16 FEV. 1971

CONTRIBUTION A LA THEORIE DE

L'ESTIMATION FONCTIONNELLE

D. BOSQ

CONSTITUTION A LA THEORIE DE

L'ESTIMATION FONCTIONNELLE



TABLE DES MATIÈRES

	Pages
INTRODUCTION	1
Chapitre I - GÉNÉRALITÉS SUR L'ESTIMATION FONCTIONNELLE	3
I - Généralités sur l'estimation statistique	3
II - Estimation de paramètres fonctionnels	5
III - Estimation sans biais	9
IV - Estimation convergente	23
V - Estimation symétrique	36
Chapitre II - ESTIMATION DE LA DENSITÉ DE PROBABILITÉ	43
I - Généralités	43
II - Convergence en moyenne quadratique intégrée	45
III - Convergence uniforme en moyenne quadratique et convergence uniforme presque sûre	51
IV - Loi limite de l'estimateur	64
V - Applications	68
Chapitre III - ESTIMATION DES DÉRIVÉES DE LA DENSITÉ	72
I - Généralités	72
II - Théorèmes asymptotiques	74
III - Etude de cas particuliers	80

INTRODUCTION

L'objet de ce travail est l'étude de certains Problèmes d'estimation statistique dans lesquels le paramètre est à valeurs dans un espace fonctionnel.

Le premier chapitre est consacré à des généralités sur l'estimation fonctionnelle : nous étudions notamment les conditions d'existence d'estimateurs sans biais. Dans le cas où de tels estimateurs existent, et où l'espace des paramètres est un espace de Hilbert réel, nous montrons l'existence d'un estimateur sans biais uniformément optimal.

Cependant on constate que, dans de nombreux Problèmes d'estimation fonctionnelle, il n'est pas possible d'obtenir un estimateur sans biais (notamment dans le cas où le paramètre est « local » en un sens que nous précisons. On doit alors se limiter à une étude asymptotique du Problème : ceci nous amène à donner le principe d'une construction d'un estimateur convergent, dans le cas où le paramètre est à valeurs dans un espace métrique.

Les chapitres II et III sont consacrés à l'estimation de la densité et de ses dérivées ; la méthode employée est celle des fonctions orthogonales. Cette méthode nous semble, en effet, plus simple et plus lisse que la méthode du noyau. Nous étudions essentiellement les propriétés asymptotiques de l'estimateur, puis nous appliquons les résultats obtenus aux estimateurs construits à partir de deux systèmes orthogonaux importants : les fonctions trigonométriques et les fonctions d'Hermite.

Dans le chapitre IV, nous montrons comment, à partir d'un estimateur de la densité, on peut construire des estimateurs pour la densité conditionnelle et les fonctions de régression.

Enfin, les deux derniers chapitres sont consacrés à des questions liées plus ou moins directement à l'estimation du support d'une loi de probabilité :

Dans le chapitre V, on généralise des résultats de M. GEFFROY sur l'estimation de la fonction frontière et, dans le chapitre VI, on étudie les éléments extrémaux d'un échantillon bivarié et on construit, à l'aide des éléments maximaux de l'échantillon, un estimateur de l'ensemble des éléments maximaux du support.

Nous ne voudrions pas terminer cette présentation sans remercier vivement Monsieur le Professeur DUGUÉ de nous avoir fait l'honneur de présider le jury de cette thèse.

Nous tenons à exprimer notre profonde gratitude à Monsieur le Professeur GEFROY pour ses encouragements et la constante bienveillance qu'il a bien voulu nous manifester ; il nous a tout particulièrement aidé en nous communiquant certains de ses résultats inédits dont il a été tiré un grand parti dans le chapitre VI.

Nous remercions également Monsieur le Professeur MALLIAVIN d'avoir bien voulu faire partie de notre jury et de nous avoir proposé un deuxième sujet.

I

GÉNÉRALITÉS SUR L'ESTIMATION FONCTIONNELLE

I - GÉNÉRALITÉS SUR L'ESTIMATION STATISTIQUE.

1. Schéma d'un problème d'estimation.

Si l'on se limite au cas d'observations indépendantes et équidistribuées, on peut schématiser un problème d'estimation statistique de la façon suivante :

Soit $\mathcal{A}$ une tribu sur un ensemble Ω et soit $\mathcal{P} = \mathcal{P}(\Omega, \mathcal{A})$ l'espace des mesures de probabilité sur l'espace mesurable $(\Omega, \mathcal{A})$.

Soit $\mathcal{D}$ un sous-espace de $\mathcal{P}$, soit Θ un ensemble, a priori quelconque, appelé espace des paramètres. Soit, enfin, φ une application de $\mathcal{D}$ dans Θ .

Le problème est d'estimer $\theta = \varphi(P)$ à partir d'observations qui proviennent d'une loi de probabilité inconnue $P \in \mathcal{D}$. L'application qui associe à ces observations la valeur estimée du paramètre s'appellera un *estimateur*. Plus précisément :

DÉFINITION 1 :

Un *estimateur d'ordre n* est une application de Ω^n dans Θ .

NOTATIONS :

Dans ce chapitre, sauf avis contraire, on désignera par T_n un estimateur d'ordre n et par $T = (T_1, \dots, T_n, \dots)$ une suite d'estimateurs.

2. Définitions.

Pour évaluer θ , on cherchera, évidemment, à obtenir des estimateurs possédant certaines propriétés de régularité et pour lesquels l'erreur

commise dans l'estimation (cette notion étant à préciser) est aussi petite que possible.

Nous donnons, ci-dessous, les définitions d'un certain nombre de propriétés que l'on peut exiger d'un estimateur.

DÉFINITION 2 :

Si Θ est muni d'une tribu $\mathcal{B}$, on dira que T_n est **mesurable** s'il est $\mathcal{Q}^n - \mathcal{B}$ mesurable.

DÉFINITION 3 :

On dira que T_n est **strict** si $T_n(\Omega^n)$ est inclus dans $\varphi(\mathcal{D})$.

DÉFINITION 4 :

Si Θ est un espace topologique muni de sa tribu Borélienne, on dira que T **converge en probabilité** si $\forall P \in \mathcal{D}$, $\forall U$ voisinage ouvert de $\varphi(P)$, on a :

$$\lim_{n \rightarrow \infty} P_n^* [T_n \in U] = 0 \quad (1)$$

P_n^* désignant la mesure extérieure associée à la mesure de probabilité P_n définie sur la tribu produit $\mathcal{Q}^n$.

DÉFINITION 5 :

Avec les mêmes hypothèses sur Θ , on dira que T **converge presque sûrement** si

$$P^\infty^* [T_n \not\rightarrow \varphi(P)] = 0 \quad \forall P \in \mathcal{D} \quad (2)$$

P^∞^* designant la mesure extérieure associée à la mesure de probabilité produit P^∞ définie sur la tribu produit $\mathcal{Q}^\infty = \mathcal{Q}^\infty$.

DÉFINITION 6 :

On suppose que Ω est un espace topologique localement compact et que Θ est un espace vectoriel topologique localement convexe séparé.

On dira que T_n est un **estimateur sans biais** de $\varphi(P)$ si, $\forall P \in \mathcal{D}$:

1) T_n est scalairement essentiellement P^n -intégrable et admet une intégrale par rapport à P^n .

$$2) E_P [T_n] = \int_{\Omega^n} T_n(\omega) dP^n(\omega) = \varphi(P) \quad (3)$$

DÉFINITION 7 :

Soit $\mathcal{T}_n$ une famille d'estimateurs d'ordre n et soit $R(T_n, P)$ une fonction de risque.

On dira que T_n^0 est **uniformément optimal** (dans la famille $\mathcal{T}_n$ et pour la fonction de risque R) si, $\forall P \in \mathcal{D}$, $\forall T_n \in \mathcal{T}_n$

$$R(T_n^0, P) \leq R(T_n, P) \quad (4)$$

Il n'est pas possible, en général, de construire un estimateur vérifiant toutes les propriétés précédentes.

Nous verrons que, sous des conditions assez larges, on peut obtenir un estimateur strict, mesurable et convergent (§IV). Au contraire, l'existence d'un estimateur sans biais n'est vérifiée que dans des cas particuliers (cf §III) ; il en est de même pour l'existence d'un estimateur uniformément optimal (cf §V).

II - ESTIMATION DE PARAMÈTRES FONCTIONNELS.

1. Exemples.

Dans la plupart des problèmes classiques d'estimation, l'espace Θ est un sous-ensemble de $\mathbf{R}$ ou de $\mathbf{R}^k$. Dans ce travail, nous étudierons principalement des problèmes d'estimation où l'espace Θ est soit un espace vectoriel topologique (en général normé), soit un espace métrique (ou pseudo-métrique).

Indiquons quelques exemples de problèmes de ce type.

Exemple 1 :

$$\Omega = \mathbf{R}^k, \mathcal{A} = \mathcal{B}_k \text{ (tribu Borélienne de } \mathbf{R}^k),$$

$\mathcal{D} = \mathcal{Y}(\mathbf{R}^k, \mathcal{B}_k), \Theta = B(\mathbf{R}^k, \mathbf{R})$ (espace des applications bornées de $\mathbf{R}^k$ dans $\mathbf{R}$ muni de la topologie de la convergence uniforme).

$\varphi(P) = F$ *fonction de répartition* de P :

$$F(x_1, \dots, x_k) = P \left[(-\infty, x_1[, (-\infty, x_2[, \dots, (-\infty, x_k[\right]$$

$$\forall (x_1, \dots, x_k) \in \mathbf{R}^k.$$

Exemple 2 :

μ mesure sur $(\Omega, \mathcal{A})$ quelconque.

$\mathcal{D} =$ Ensemble des lois de probabilités sur $(\Omega, \mathcal{A})$ admettant, par rapport à μ , une densité de Radon-Nikodym de carré μ -intégrable.

$\Theta =$ espace de Hilbert $L^2(\mu)$

$$\varphi(P) = \frac{dP}{d\mu}.$$

Exemple 3 :

$$\Omega = \mathbf{R}, \mathcal{A} = \mathcal{B} \text{ (tribu Borélienne de } \mathbf{R})^{(1)},$$

$\mathcal{D} =$ ensemble des lois de probabilité P absolument continues par rapport à la mesure de Lebesgue λ sur $\mathbf{R}$ et possédant une version continue de la *densité* : f_p .

$\Theta = C(\mathbf{R})$ (Espace des fonctions continues bornées sur $\mathbf{R}$ muni de la topologie de la convergence uniforme).

$$\varphi(P) = f_p.$$

(1) Dans la suite, nous supposons, sauf avis contraire, que les espaces topologiques sont munis de leur tribu Borélienne.

Exemple 4 :

$\Omega = \mathbf{R}^2$, $\mathcal{D}$ = Ensemble des mesures de probabilité absolument continues, à densité $f(x, y)$ continue, telles que

$$\int_{\mathbf{R}} y f(x, y) dy$$

existe et soit continue $\forall x \in \mathbf{R}$ et telles que

$$\int_{\mathbf{R}} f(x, y) dy$$

soit continue et positive $\forall x \in \mathbf{R}$.

Θ = espace des fonctions continues sur $\mathbf{R}$, muni de la topologie de la convergence uniforme sur tout compact.

$\varphi(P)$ = **Fonction de régression** de y en x =

$$\int_{\mathbf{R}} y f(x, y) dy / \int_{\mathbf{R}} f(x, y) dy$$

Exemple 5 :

$\Omega = \mathbf{R}^2$, $\mathcal{D}$ = Ensemble des mesures de probabilité absolument continues, dont le support S est de la forme :

$$S = \{ (x, y) \in \mathbf{R}^2 \mid x \in [0, 1], 0 \leq y \leq s_p(x) \}$$

où s_p est une fonction définie et continue sur $[0, 1]$.

$\Theta = C[0, 1]$ (espace des fonctions continues sur $[0, 1]$ muni de la topologie de la convergence uniforme),

$\varphi(P) = s_p$ (on dira que s_p est la **fonction frontière** de P).

Exemple 6 :

$\Omega = \mathbf{R}^2$, $\mathcal{D}$ = ensemble des lois de probabilité absolument continues, à

support compact.

Θ = Ensemble des compacts de $\mathbf{R}^2$ muni de la topologie induite par la pseudo-métrique :

$$d(A,B) = \lambda^2 (A \Delta B) \quad \forall A, B \in \Theta$$

où λ^2 désigne la mesure de Lebesgue sur $\mathbf{R}^2$ et Δ l'opération différence symétrique.

$\varphi(P)$ = Support de P .

Nous reviendrons sur ces exemples dans la suite de ce travail.

2. Choix d'une fonction de coût.

Pour comparer l'efficacité des diverses méthodes d'estimation, il est nécessaire de choisir une fonction de coût.

Si Θ est un espace normé, il semble naturel de choisir la fonction

$$\begin{aligned} 1 \left[T_n(\omega_1, \dots, \omega_n) ; \varphi(P) \right] &= || T_n(\omega_1, \dots, \omega_n) - \varphi(P) ||^2 \\ \forall (\omega_1, \dots, \omega_n) &\in \Omega^n, \forall P \in \mathcal{D} \end{aligned} \quad (5)$$

La fonction de risque correspondante s'écrit alors :

$$R(T_n, \varphi(P)) = E_P || T_n(\omega_1, \dots, \omega_n) - \varphi(P) ||^2 \quad \forall P \in \mathcal{D} \quad (6)$$

Elle généralise le MISE (Mean integrated square error) introduit par PARZEN [1] ; en effet, le MISE est défini par :

$$J_n = E \int_{\mathbf{R}} [\hat{f}_n(x) - f(x)]^2 dx \quad f \in L^2, \hat{f}_n \in L^2 \quad (7)$$

où f est le paramètre ($\Theta = L^2$) et $\hat{f}_n$ l'estimation de f .

Propriétés de ℓ

Si Θ vérifie les hypothèses convenables, la fonction $\ell(q) = ||q||^2$

a les propriétés usuelles d'une fonction de coût :

$$1) \quad q = 0 \implies \ell(q) = 0 \quad (8)$$

$$2) \quad \ell(-q) = \ell(q) \quad (9)$$

3) Dans un treillis vectoriel normé :

$$|q_2| \succ |q_1| \implies \ell(|q_2|) \succ \ell(|q_1|) \quad (10)$$

4) Dans un espace de Hilbert :

$$\ell(\alpha q_1 + (1-\alpha) q_2) \leq \alpha \ell(q_1) + (1-\alpha) \ell(q_2) \quad \forall \alpha \in [0,1] \quad (11)$$

REMARQUES.

a) La définition de ℓ s'étend sans difficulté au cas où Θ est un espace pseudo-métrique.

b) La fonction ℓ a l'avantage d'être maniable (surtout dans un espace de Hilbert), mais elle ne donne pas une idée entièrement satisfaisante de la précision d'une estimation. Dans L^2 (par exemple), ℓ ne tient pas compte de la corrélation entre $\hat{f}_x(x)$ et $\hat{f}_n(x')$ ($x \neq x'$). Toutefois, il semble difficile de définir une fonction de coût plus satisfaisante que ℓ , et qui reste facilement utilisable dans la pratique.

III - ESTIMATION SANS BIAIS.

Dans ce paragraphe, nous donnons quelques propriétés des estimateurs sans biais et étudions des conditions d'existence pour de tels estimateurs.

1. Propriétés d'un estimateur sans biais.

a) Rappelons que si T_n est scalairement essentiellement P^n -intégrable et admet une intégrale par rapport à P^n , cette intégrale est caractérisées par la relation :

$$\langle \int_{\Omega^n} T_n(\omega) dP^n(\omega), \theta' \rangle = \int_{\Omega^n} \langle T_n(\omega), \theta' \rangle dP^n(\omega) \quad (12)$$

$\forall \theta' \in \Theta'$ dual topologique de Θ . Compte tenu de la définition 6 (I - 2), le fait que T_n est sans biais s'écrit donc :

$$\int_{\Omega^n} \langle T_n(\omega), \theta' \rangle dP^n(\omega) = \langle \varphi(P), \theta' \rangle \quad \forall \theta' \in \Theta', \quad \forall P \in \mathcal{D} \quad (13)$$

b) Donnons, maintenant, une généralisation d'une propriété bien connue de la variance d'un estimateur sans biais.

THÉORÈME 1 :

Si Θ est un espace de Hilbert réel, si T_n est essentiellement P^n -intégrable et si $E_P(T_n) \in \varphi(\mathcal{D})$, $\forall P \in \mathcal{D}$. alors, pour que T_n soit sans biais, il faut et il suffit que :

$$E_P \|T_n - \varphi(P)\|^2 < E_P \|T_n - \theta\|^2 \quad \forall \theta \in \varphi(\mathcal{D}), \quad \forall P \in \mathcal{D} \quad (14)$$

DÉMONSTRATION.

Si T_n est essentiellement P^n -intégrable, on sait (cf, COURREGÉ [2] p. 227 th. 1A) que T_n est scalairement essentiellement P^n -intégrable et qu'il admet une intégrale par rapport à P^n .

D'autre part, il est facile de voir que $\|T_n - \theta\|^2$ est P^n -mesurable, $\forall \theta \in \varphi(\mathcal{D})$ (cf, COURREGÉ [2] p. 227 prop. 4) donc que les deux membres de (14) ont un sens (leurs valeurs étant finies ou infinies).

Ceci dit, on peut écrire, $\forall \theta \in \varphi(\mathcal{D})$:

$$\begin{aligned} \|T_n - \theta\|^2 &= \|T_n - E_P(T_n)\|^2 + \|E_P(T_n) - \theta\|^2 + 2 \langle T_n - E_P(T_n), E_P(T_n) - \theta \rangle \\ &\quad \dots E_P(T_n) - \theta \rangle \end{aligned} \quad (15)$$

En écrivant (15) pour $\theta = \varphi(P)$ et en faisant la différence, il vient :

$$\begin{aligned} ||T_n - \theta||^2 - ||T_n - \varphi(P)||^2 &= ||E_P(T_n) - \theta||^2 - ||E_P(T_n) - \varphi(P)||^2 \\ &+ 2 \langle T_n - E_P(T_n), \varphi(P) - \theta \rangle \end{aligned} \quad (16)$$

D'où, en passant aux espérances et en tenant compte de (12) :

$$\begin{aligned} E_P ||T_n - \theta||^2 - E_P ||T_n - \varphi(P)||^2 &= ||E_P(T_n) - \theta||^2 - ||E_P(T_n) - \varphi(P)||^2 \\ &\dots - \varphi(P)||^2 \end{aligned} \quad (17)$$

$$\text{Donc (14)} \Leftrightarrow ||E_P(T_n) - \theta||^2 \geq ||E_P(T_n) - \varphi(P)||^2$$

$$\forall \theta \in \varphi(\mathcal{D}), \forall P \in \mathcal{P} \quad (18)$$

Si T_n est sans biais, on a (18) donc (14) ; réciproquement, si (14) est vérifié, on a (18) et en prenant $\theta = E_P(T_n)$ dans (18), on en déduit que T_n est sans biais.

REMARQUES :

a) si $\varphi(\mathcal{D})$ est un cône convexe complet, si T_n est strict et si l'application $\omega \rightarrow T_n(\omega)$ est continue à support compact, alors T_n est P^n -intégrable et $E_P(T_n) \in \varphi(\mathcal{D}), \forall P \in \mathcal{P}$. (cf. COURREGES [2] p.216, cor. 1).

b) LEHMANN [3] a proposé une définition de la notion d'estimateur sans biais ⁽¹⁾ valable pour un Θ quelconque : R désignant une fonction de risque, on dit que T_n est sans biais si :

$$R(T_n, \varphi(P)) \leq R(T_n, \theta) \quad \forall \theta \in \varphi(\mathcal{D}) \quad (19)$$

Le théorème 1 montre que si R désigne la fonction de risque définie par (6) et si Θ est un espace de Hilbert, la définition 6 coïncide avec celle de Lehmann.

(1) et plus généralement de la notion de procédure statistique sans biais.

2. Cas d'un paramètre fonctionnel.

Les paramètres que nous voulons estimer étant en général des fonctions, il est intéressant de savoir dans quelle mesure un estimateur sans biais « point par point » est sans biais au sens de la définition 6.

Plus précisément, si Θ est un espace de fonctions $x \rightarrow f(x)$ définies sur un ensemble E quelconque et à valeurs réelles, et si l'on note $T_n(\omega, x)$ un estimateur de $\varphi(P)$, est-ce que l'on a l'équivalence

$$T_n \text{ sans biais} \Leftrightarrow E_P(T_n(\omega, x)) = \varphi(P)(x), \forall x \in E ? \quad (20)$$

Nous allons voir que la réponse est positive pour les espaces L^2 et $C[0,1]$, des hypothèses convenables étant faites sur T_n .

LEMME 1 :

si $\Omega = \mathbf{R}$, $\Theta = L^2$ et si $T_n(\omega, x)$ est de carré P^n x λ -intégrable alors, T_n est sans biais si et seulement si

$$E_P[T_n^*(\omega, x)] = \varphi(P)^*(x) \quad \text{p.p.}^{(1)} \quad (21)$$

DÉMONSTRATION :

T_n étant toujours supposé scalairement essentiellement P^n intégrable et intégrable dans L^2 , on a, d'après (12) :

$$\int_{\mathbf{R}} \Psi(P)^*(x) g^*(x) dx = \int_{\mathbf{R}} dP^n(\omega) \int_{\mathbf{R}} T_n^*(\omega, x) g^*(x) dx \quad (22)$$

$$\forall g \in L^2$$

$$\text{où l'on a posé} \quad \Psi(P) = \int_{\Omega^n} T_n(\omega) dP^n(\omega)$$

(1) si $f \in L^2$, f^* désigne une version de f .

en appliquant le théorème de Fubini, il vient, en supposant (21) vérifié :

$$\int_{\mathbf{R}} \Psi(P)^*(x) \cdot g^*(x) dx = \int_{\mathbf{R}} g^*(x) dx \int_{\mathbf{R}^n} T_n^*(\omega, x) dP_n(\omega) \quad (23)$$

$$= \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (24)$$

on en tire $\Psi(P)^*(x) = \varphi(P)^*(x)$ p.p. donc T_n est sans biais.

Réciproquement : si T_n est sans biais, on peut écrire :

$$\int_{\Omega^n} dP^n(\omega) \int_{\mathbf{R}} T_n^*(\omega, x) g(x) dx = \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (25)$$

d'où, d'après le théorème de Fubini :

$$\int_{\mathbf{R}} g^*(x) dx \int_{\Omega^n} T_n^*(\omega, x) dP^n(\omega) = \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (26)$$

$$\text{donc} \quad \int_{\Omega} T_n^*(\omega, x) dP^n(\omega) = \varphi(P)^*(x) \quad \text{p.p.} \quad (27)$$

C.Q.F.D.

LEMME 2 :

Si $\Omega = \mathbf{R}$, $\Theta = \mathbf{C} [0,1]$, si $T_n(\omega, x)$ est $P^n_{x\mu}$ -intégrable, $\forall \mu$ mesure sur $[0,1]$ et si $\int_{\mathbf{R}^n} T_n(\omega, x) dP_n(\omega)$ existe $\forall x \in [0,1]$, $\forall P \in \mathcal{D}$; alors, T_n est sans biais si et seulement si :

$$E_P [T_n(\omega, x)] = \varphi(P)(x) \quad \forall x \in [0,1] \quad (28)$$

DÉMONSTRATION : T_n étant encore supposé scalairement essentiellement P^n -intégrable et intégrable dans $\mathbf{C} [0,1]$, $\forall P \in \mathcal{D}$; la

démonstration est analogue à celle du lemme 1 (utilisation de (12) et du théorème de Fubini).

Ces deux lemmes ramènent, dans les espaces de fonctions usuels, la recherche d'estimateurs sans biais à la recherche d'estimateurs sans biais « point par point ».

REMARQUE :

La situation est plus compliquée dans $C(\mathbf{R})$ car Θ' est alors l'espace vectoriel des fonctions d'ensembles additives, régulières et bornées, définies sur l'Algèbre engendrée par les fermés de $\mathbf{R}^{(1)}$.

La démonstration du lemme 2 permet seulement de montrer que (28) est une condition nécessaire pour que T_n soit sans biais.

3. Existence d'un estimateur sans biais.

Les conditions d'existence d'un estimateur sans biais ont été étudiées principalement par HALMOS [5] et par BICKEL et LEHMANN [6]. Nous complétons leurs résultats en étudiant l'estimation sans biais de paramètres que nous appellerons « paramètres locaux ».

α) LE THÉORÈME DE BICKEL-LEHMANN.

DÉFINITION 8 :

On dit que φ est estimable s'il existe un entier n pour lequel φ admet un estimateur sans biais d'ordre n .

Si φ est estimable, le degré de φ est le plus petit entier n pour lequel φ est estimable.

EXEMPLES :

a) $\Omega = \mathbf{R}$, $\mathcal{D}$ = ensemble des mesures de probabilité sur $(\mathbf{R}, \mathcal{B})$ à variance finie, $\varphi(P) = \text{variance de } P$: φ est estimable de degré 2.

(1) cf. DUNFORD-SCHWARTZ [4] 1 p. 261-62.

b) Mêmes hypothèses, $\varphi(P)$ = écart type de P : φ n'est pas estimable.

c) Exemple 1 du §II : la **fonction de répartition** est estimable de degré 1, point par point.

d) $\Omega = \mathbf{R}$, $\mathcal{D} = \mathcal{P}(\mathbf{R})$, $\Theta = \mathbf{C}(\mathbf{R}, \mathbf{C})$ (espace des applications continues bornées de $\mathbf{R}$ dans $\mathbf{C}$), $\varphi(P) = \int_{\mathbf{R}} e^{itx} dP(x)$ $t \in \mathbf{R}$: la **fonction caractéristique** est estimable de degré 1, point par point.

THÉORÈME 2 (BICKEL-LEHMANN).

Pour $\Theta = \mathbf{R}^{(1)}$, si $\mathcal{D}$ est convexe et si φ est estimable, pour que φ soit de degré n , il faut et il suffit que :

1°) $\forall P, Q \in \mathcal{D}$ $\varphi[\alpha P + (1-\alpha)Q]$ est un polynôme de degré $\leq n$ en α ($\alpha \in [0,1]$).

2°) $\forall P \in \mathcal{D}$, $\exists Q \in \mathcal{D}$ tel que $\varphi[\alpha P + (1-\alpha)Q]$ est un polynôme de degré n en α ($\alpha \in [0,1]$).

DÉMONSTRATION.

Voir BICKEL-LEHMANN [6] p, 1525-26.

APPLICATION :

Il est clair que, si $\varphi[\alpha P + (1-\alpha)Q]$ n'est pas un polynôme, le théorème 2 permet d'affirmer que φ n'est pas estimable.

EXEMPLES :

e) Si on reprend l'exemple 4 (II - 1), on voit immédiatement que $\mathcal{D}$ est convexe et que φ n'est pas un polynôme : la **fonction de régression n'est pas estimable point par point**.

(1) Ce théorème se généralise immédiatement au cas où Θ est un espace de Banach.

f) En reprenant les hypothèses de l'exemple 4 (II - 1) et en posant $\varphi(P) = f(x,y) / \int_{\mathbf{R}} f(x,y) dy$ (densité de y conditionnée par x), on en déduit **que la densité conditionnelle n'est pas estimable point par point.**

g) Si on reprend l'exemple 5 (II - 1), on vérifie de même que $\mathcal{D}$ est convexe et que φ n'est pas un polynôme d'où, en appliquant le lemme 2 (III - 2) ⁽¹⁾ :

la fonction frontière n'est pas estimable.

h) $\Omega = \mathbf{R}$, $\mathcal{D} = \mathcal{P}(\mathbf{R})$, Θ = Ensemble des applications bornées de $\mathbf{R}^+$ dans $\mathbf{R}$ muni de la topologie de la convergence uniforme, $\varphi(P) = q$ avec

$q(\ell) = \sup_{z \in \mathbf{R}} P([z, z+\ell])$, $\forall \ell \geq 0$: **la fonction de concentration n'est pas**

estimable point par point. (On vérifie facilement que $\varphi(\alpha \delta_a + (1-\alpha) \delta_b)$, où δ_a et δ_b sont les mesures de Dirac aux points a et b ($a \neq b$), n'est pas un polynôme en α).

β) CAS D'UN PARAMÈTRE LOCAL.

Dans ce paragraphe, on suppose que Ω est un espace topologique localement compact métrisable et que μ est une mesure σ -finie sur Ω .

DÉFINITION 9 :

Soit F un fermé de Ω , on dira que φ est localisé par F si, $\forall U$ voisinage ouvert de F ; $\forall P, Q \in \mathcal{D}$,

$$P \upharpoonright U = Q \upharpoonright U \implies \varphi(P) = \varphi(Q)$$

$P \upharpoonright U$ et $Q \upharpoonright U$ désignant les restrictions de P et Q à U .

On dira que φ est local s'il est localisé par un point.

On dira que φ est μ -local s'il est localisé par un fermé de μ -mesure nulle.

(1) Ou directement le théorème 2.

Enfin, on dira que φ est global s'il n'est localisé que par $F = \Omega$.

EXEMPLES :

i) $\Omega = \Theta = \mathbf{R}$, $\mathcal{D}$ = ensemble des mesures de probabilité à fonction de répartition strictement croissante, $\varphi(P)$ = médiane de P , φ est un paramètre global.

j) $\Omega = \Theta = \mathbf{R}$, $\mathcal{D} = \mathcal{P}$, $\varphi(P) = P(-\infty, a]$, a donné $\in \mathbf{R}$. φ est localisé par $(-\infty, a]$. En revanche, la fonction de répartition (cf, exemple 1 (II - 1)) est évidemment un paramètre global.

k) $\Omega = \Theta = \mathbf{R}$, $\mathcal{D} = \mathcal{P}$, $\varphi(P) = P(\{a\})$, a donné $\in \mathbf{R}$. φ est local, Si μ est la mesure de Lebesgue sur $\mathbf{R}$, φ est aussi μ -local.

DÉFINITION 10 :

Θ étant un espace vectoriel, on dira que φ est un paramètre **homogène** sur F (fermé de Ω) si, $\forall U$ voisinage ouvert de F et $\forall P \in \mathcal{D}$:

$$1^\circ) P(U) > 0 \Rightarrow P_U \in \mathcal{D}^{(1)} \text{ et } \varphi(P_U) = \frac{\varphi(P)}{P(U)} \quad (30)$$

$$2^\circ) P(U) = 0 \Rightarrow \varphi(P) = 0$$

EXEMPLES :

1) $\Omega = \mathbf{R}^p$, $\Theta = \mathbf{R}$, $\mathcal{D}$ = ensemble des mesures de probabilité P absolument continues par rapport à λ^p et à densité f_P continue dans un voisinage du point a , a donné $\in \mathbf{R}$.

$\varphi(P) = f_P(a)$. φ est un paramètre homogène sur $\{a\}$.

m) $\Omega = \Theta = \mathbf{R}$, $\mathcal{D}$ = ensemble des mesures de probabilité P absolument continues, à densité f_P k fois continuellement différentiable dans un voisinage du point $a \in \mathbf{R}$.

$\varphi(P) = f_P^{(k)}(a)$. φ est un paramètre homogène sur $\{a\}$.

(1) P_U désigne la mesure de probabilité induite par P sur U .

- Remarquons *qu'un paramètre homogène sur F est localisé par F*. En effet :

$$P \upharpoonright U = Q \upharpoonright U \implies P(U) = Q(U)$$

si $P(U) > 0$, on a aussi $P_U = Q_U$, donc $\varphi(P_U) = \varphi(Q_U)$

d'où $\varphi(P) = \varphi(Q)$. Si $P(U) = 0$, on a immédiatement $\varphi(P) = \varphi(Q) = 0$;

La réciproque n'est pas vraie : la fonction frontière au point a ($a \in]0,1[$) est localisée par la droite $x = a$ mais n'est pas homogène sur cette droite.

Donnons maintenant deux résultats relatifs à l'estimation sans biais des paramètres locaux.

THÉORÈME 3 :

Si φ est un paramètre à valeurs réelles, non constant, homogène sur un fermé F et tel que, $\forall P \in \mathcal{D}$, F soit de P-mesure nulle, alors φ est de degré supérieur à 1⁽¹⁾.

DÉMONSTRATION.

Nous allons raisonner par l'absurde.

Soit T un estimateur de φ , sans biais d'ordre 1, on a :

$$\int_{\Omega} T(\omega) dP(\omega) = \varphi(P) \quad \forall P \in \mathcal{D} \quad (32)$$

Soit $(U_k)_{k \in \mathbb{N}}$ une suite décroissante de voisinages ouverts de F, telle que $U_k \downarrow F$ quand $k \rightarrow \infty$.

Si $P(U_k) > 0$, on peut écrire, compte tenu de (30) :

$$\varphi(P_{U_k}) = \int_{\Omega} T(\omega) dP_{U_k}(\omega) = \int_{U_k} T(\omega) \frac{dP(\omega)}{P(U_k)} = \frac{\varphi(P)}{P(U_k)} \quad (33)$$

(1) Si φ n'est pas estimable, nous posons par convention degré de $\varphi = +\infty$.

On en tire :

$$\int_{U_k} T(\omega) dP(\omega) = \varphi(P) \quad (34)$$

si $P(U_k) = 0$, on a, d'après (31) :

$$\int_{U_k} T(\omega) dP(\omega) = \varphi(P) = 0. \quad (35)$$

Donc, en tenant compte de (32), on peut écrire dans tous les cas :

$$\int_{U_k^c} T(\omega) dP(\omega) = 0 \quad (36)$$

On peut donc construire une suite (T_k) d'estimateurs de φ sans biais d'ordre 1 en posant :

$$T_k(\omega) = T(\omega) I_{U_k}(\omega) \quad \forall \omega \in \Omega \quad (37)$$

$$\forall k \in \mathbb{N}$$

Quand $k \rightarrow \infty$, $T_k(\omega) \rightarrow T(\omega) I_F(\omega) = T_0(\omega) \quad \forall \omega \in \Omega$

d'autre part. $|T_k(\omega)| \leq |T(\omega)|, \quad \forall \omega \in \Omega$

on peut donc appliquer le théorème de Lebesgue et écrire :

$$\int_{\Omega} T_k(\omega) dP(\omega) \rightarrow \int_{\Omega} T_0(\omega) dP(\omega)$$

or

$$\int_{\Omega} T_k(\omega) dP(\omega) = \varphi(P) \quad \forall P \in \mathcal{D}$$

(1) $I_{U_k}(\omega)$ désigne l'indicateur de U_k .

20

et

$$\int_{\Omega} T_o(\omega) dP(\omega) = \int_F T_o(\omega) dP(\omega) = 0 \quad \forall P \in \mathcal{D}$$

d'où :

$$\varphi(P) = 0 \quad \forall P \in \mathcal{F}$$

ce qui est contradictoire. φ est donc bien de degré supérieur à 1.

COROLLAIRE 1 :

Si φ vérifie les hypothèses du théorème 3, si $\mathcal{D}$ est convexe et si $\varphi[\alpha P + (1 - \alpha) Q]$ est un polynôme de degré 1 en α , alors φ n'est pas estimable.

C'est une conséquence des théorèmes 2 et 3.

APPLICATION :

On peut appliquer le corollaire 1 aux paramètres des exemples 1) et m), on obtient donc les résultats suivants :

- 1) *La densité multivariée n'est pas estimable point par point.*
- 2) *Les dérivées de la densité ne sont pas estimables point par point.*

REMARQUE :

Le fait que la densité univariée n'admet pas d'estimateur sans biais a déjà été démontré par ROSENBLATT [7] et par BICKEL et LEHMANN [6]. Les résultats ci-dessus sont plus généraux et, nous semble-t-il, ont été obtenus d'une manière plus naturelle.

Démontrons, pour terminer, le

THÉORÈME 4 :

Soit φ un paramètre à valeurs réelles, non constant, μ -local et tel que

1°) $P \in \mathcal{D} \Rightarrow P$ est absolument continue par rapport à μ

2°) Si F est un fermé qui localise φ , si U est un voisinage ouvert de F et si P est tel que $0 < P(U) < 1$ alors, la famille $\mathcal{D}(U, P)$ de mesures de probabilité $Q \in \mathcal{D}$ telles que $Q_U = P_U$ est μ -complète ⁽¹⁾.

3°) $P(U) = 0 \Rightarrow \varphi(P) = 0$ quel que soit U voisinage ouvert de F .

Dans ces conditions, φ est de degré supérieur à 1.

DÉMONSTRATION :

Comme pour le théorème 3, nous allons procéder par l'absurde :

Soit (U_k) une suite d'ouverts de Ω , décroissante et tendant vers F et soit $P \in \mathcal{D}$. On aura toujours, au moins pour k assez grand, $P(U_k) < 1$. En effet, si l'on avait $P(U_k) = 1$, $\forall k \in \mathbb{N}$, on en déduirait $P(F) = 1$, ce qui est contradictoire avec le fait que P est absolument continue par rapport à μ et que $\mu(F) = 0$;

D'autre part, on sait que si $P(U_k) = 0$, $\varphi(P) = 0$; il suffit donc de montrer qu'il en est de même si $0 < P(U_k) < 1$ pour obtenir la contradiction cherchée.

Soit T un estimateur de φ , sans biais, d'ordre 1; soit $Q \in \mathcal{D}(U_k, P)$. On a les deux égalités :

$$\int_{U_k} T(\omega) dP(\omega) + \int_{U_k^c} T(\omega) dP(\omega) = \varphi(P) \quad (38)$$

et

$$\int_{U_k} T(\omega) dP(\omega) + \int_{U_k^c} T(\omega) dQ(\omega) = \varphi(Q) \quad (39)$$

(1) Autrement dit : $\int_{\Omega} T(\omega) dQ(\omega) = 0$, $\forall Q \in \mathcal{D}(U, P) \Rightarrow T(\omega) = 0$
 μ -p.p.

Comme φ est μ -local, on en tire :

$$\int_{U_k^c} T(\omega) dQ(\omega) = \int_{U_k^c} T(\omega) dP(\omega) = \alpha(P) \quad \forall Q \in \mathcal{D}(U_k, P) \quad (40)$$

Posons $T_1(\omega) = T(\omega) I_{U_k^c}(\omega)$, il vient :

$$\int_{\Omega} [T_1(\omega) - \alpha(P)] dQ(\omega) = 0 \quad \forall Q \in \mathcal{D}(U_k, P) \quad (41)$$

d'où, comme $\mathcal{I}(U_k, P)$ est μ -complète :

$$T_1(\omega) - \alpha(P) = 0 \quad \mu \text{ p.p.}$$

Or, comme $P(U_k) > 0$, on a aussi $\mu(U_k) > 0$, d'où

$$T_1(\omega) - \alpha(P) = 0 \quad \mu \text{ p.p. sur } U_k$$

c'est-à-dire que $\alpha(P) = 0$.

On en conclut que, $\forall P \in \mathcal{D}$ tel que $0 < P(U_k) < 1$

on a

$$\int_{U_k^c} T(\omega) dP(\omega) = 0$$

Comme $U_k^c \uparrow \Omega - F$, on en tire

$$\varphi(P) = \int_{\Omega} T(\omega) dP(\omega) = \int_{\Omega - F} T(\omega) dP(\omega) = 0 \quad \text{CQFD}$$

COROLLAIRE 2 :

Si φ vérifie les hypothèses du théorème 4, si $\mathcal{D}$ est convexe et si $\varphi[\alpha P + (1 - \alpha) Q]$ est un polynôme de degré 1 en α , φ n'est pas estimable.

C'est une conséquence des théorèmes 2 et 4 .

γ) CONCLUSION

On voit que, dans de nombreux problèmes d'estimation de fonctions, il n'est pas possible d'obtenir un estimateur sans biais. On peut alors essayer de remplacer la condition (3) par une condition moins forte. Généralisant et modifiant légèrement la notion de « unbiased touch estimate » proposée par TUKEY [8], on peut donner la définition suivante :

DEFINITION 6 bis :

Soit $f(x)$ une fonction continue et soit $T_n(\omega, x)$ un estimateur de $f(x)$.
On dira que T_n est ϵ -sans biais si

$$\forall x \in \mathbb{R}, \exists x' \in [x - \epsilon, x + \epsilon] \text{ tel que } E_p[T_n(\omega, x')] = f(x') .$$

De nombreux paramètres fonctionnels sont ϵ -sans biais estimables. Par exemple

$$\frac{F_n(x + \epsilon) - F_n(x - \epsilon)^{(1)}}{\epsilon}$$

est un estimateur ϵ -sans biais de la densité si celle-ci est continue.

IV - ESTIMATION CONVERGENTE.

Ce paragraphe est consacré à l'étude de l'existence et de la construction d'un estimateur convergent.

Nous procéderons par étapes en commençant par construire un estimateur de $P(P \in \mathcal{D})$, puis en en déduisant un estimateur de $\phi(P)$.

(1) F_n désigne ici la fonction de répartition empirique.

1. Estimation de P.

a) Préliminaires.

Nous supposons dans la suite, sauf avis contraire, que Ω est un espace topologique localement compact à base dénombrable.

Soit $\mathcal{M}$ l'espace des mesures de Radon sur Ω , muni de la topologie vague. On sait qu'on peut identifier l'espace $\mathcal{P}$ des mesures de probabilité sur Ω à un sous-espace topologique de $\mathcal{M}^{(1)}$. De plus, $\mathcal{P}$ est métrisable et séparable pour la topologie induite par la topologie vague⁽²⁾.

Commençons par donner un résultat préliminaire connu mais qui nous sera utile dans la suite.

LEMME 3 :

Soit $P_n = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i)^{(3)}$ la mesure empirique associée à un échantillon de taille n tiré suivant $P \in \mathcal{P}$. Soit V_0 un voisinage fondamental de l'origine dans $\mathcal{M}$ muni de la topologie vague. On a l'inégalité :

$$P^\infty [P_n - P \in V_0, \forall n \geq N] \geq 1 - \frac{C}{N} \quad \forall P \in \mathcal{P} \quad (43)$$

où C est une constante positive qui ne dépend que de V_0 .

DÉMONSTRATION.

On sait que V_0 est de la forme :

$$V_0 = \{ \mu \in \mathcal{M} \mid |\mu(\varphi_i)| \leq \varepsilon, \quad 1 \leq i \leq k \}$$

où $\varphi_1, \dots, \varphi_k \in K(\Omega)$ (espace des fonctions continues à support compact définies sur Ω) et où ε est un nombre réel positif.

(1) cf. COURREGÉ [2] théorème 4 p. 119.

(2) cf. COURREGÉ [2] exercice 3 p. 129.

(3) $\delta(\omega_i)$ désigne la mesure de Dirac au point ω_i .

L'application $\omega \rightarrow \delta(\omega)$ étant continue, on en déduit que l'événement $\{P_n - P \in V_0, \forall n \geq N\}$ est mesurable. $\forall P \in \mathcal{P}, \forall N \in \mathbb{N}$.

Soit alors M un majorant commun aux $|\varphi_i|$. D'après l'inégalité de Kolmogoroff, on peut écrire⁽¹⁾ :

$$P^\infty \left[\max_{2^h \leq n < 2^{h+1}} |P_n(\varphi_i) - P(\varphi_i)| > \varepsilon \right] \leq \frac{2^{h+1} \cdot M^2}{2^{2h} \cdot \varepsilon^2} \quad i = 1, \dots, k$$

d'où

$$P^\infty \left[\max_{n \geq 2^h} |P_n(\varphi_i) - P(\varphi_i)| > \varepsilon \right] \leq \frac{M^2}{2^{h-2} \cdot \varepsilon^2} \quad i = 1, \dots, k$$

le résultat s'en déduit aisément en posant $C = \frac{8 k M^2}{2}$

et en choisissant h tel que $2^h \leq N < 2^{h+1}$.

REMARQUE :

L'inégalité (43) implique la convergence presque sûre de P_n vers P , uniformément sur $\mathcal{P}$, c'est-à-dire que

$$\sup_{P \in \mathcal{P}} P^\infty \left[\bigcup_{n > N} P_n - P \notin V_0 \right] \rightarrow 0 \quad \text{avec} \quad \frac{1}{N}$$

$\forall V_0$ voisinage fondamental de l'origine.

b) Construction d'un estimateur strict.

Soit d une distance sur $\mathcal{M}^+$ compatible avec la topologie vague. $\mathcal{P}$ étant séparable, on peut se donner une suite de recouvrements de

(1) cf. DUGUE [9] p. 85.

$P_n(\Omega^n)$ par un nombre dénombrable de boules ouvertes de même rayon r_n tel que $\lim_{n \rightarrow \infty} r_n = 0$.

Désignons par $(B_{in})_{i \in \mathbb{N}}$ une telle suite et posons

$$d(B_{in} \cap P_n(\Omega^n), \mathcal{D}) = d_{in} \quad \forall i, n \in \mathbb{N} \quad (44)$$

Soit alors (γ_n) une suite de nombres positifs telle que $\lim_{n \rightarrow \infty} \gamma_n = 0$ et soit (P_{in}) une suite double d'éléments de $\mathcal{D}$ telle que :

$$d(B_{in} \cap P_n(\Omega^n), P_{in}) < d_{in} + \gamma_n \quad \forall i, n \in \mathbb{N} \quad (45)$$

On définira l'estimateur P_n^* de la façon suivante :

Soit $(\omega_1, \dots, \omega_n) \in \Omega^n$, il existe $i \in \mathbb{N}$ tel que

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i) \in B_{in}.$$

Parmi tous les B_{in} contenant P_n soit B_{jn} celui qui a le plus petit indice, on posera :

$$P_n^*(\omega_1, \dots, \omega_n) = P_{jn} \quad \forall (\omega_1, \dots, \omega_n) \in \Omega^n, \forall n \in \mathbb{N} \quad (46)$$

c) Convergence de P_n^* .

THÉORÈME 5 :

P_n^* est un estimateur de P strict, mesurable, et qui converge presque sûrement, uniformément sur $\mathcal{D}$.

DÉMONSTRATION.

1°) P_n^* est strict puisque $P_{in} \in \mathcal{D} \quad \forall i, n \in \mathbb{N}$.

2°) Montrons que P_n^* est mesurable : pour cela il suffit de montrer que les ensembles

$$A_{jn} = \{ \omega \in \Omega^n \mid P_n^*(\omega) = P_{jn} \}$$

sont mesurables $\forall j, n \in \mathbb{N}$.

Or, si l'on pose

$$D_{jn} = \{ \omega \in \Omega^n \mid P_n \in B_{jn} \} \quad \forall j, n \in \mathbb{N}$$

on a, en tenant compte de (46) :

$$A_{jn} = D_{1n}^c \cap D_{2n}^c \cap \dots \cap D_{j-1,n}^c \cap D_{jn} \quad (47)$$

et D_{jn} étant mesurable ($\forall j, n \in \mathbb{N}$). A_{jn} l'est aussi. CQFD.

3°) Montrons enfin la convergence de P_n .

D'après (45), il existe $Q_{jn} \in B_{jn}$ $P_n(\Omega^n)$ telle que :

$$d(Q_{jn}, P_{jn}) < d_{jn} + \gamma_n \quad \forall j, n \in \mathbb{N} \quad (48)$$

d'autre part

$$d_{jn} \leq d(P_n, P) \quad \forall j, n \in \mathbb{N} \quad (49)$$

Or

$$d(P_{jn}, P) \leq d(P_{jn}, Q_{jn}) + d(Q_{jn}, P_n) + d(P_n, P) \quad (50)$$

$$\forall j, n \in \mathbb{N}$$

Donc

$$d(P_{jn}, P) \leq \gamma_n + 2r_n + 2d(P_n, P) \quad \forall j, n \in \mathbb{N} \quad (51)$$

c'est-à-dire que, $\forall \alpha > 0$, $\forall \omega \in \Omega^n$, et pour n assez grand :

$$d(P_n(\omega), P) < \frac{\alpha}{3} \Rightarrow d(P_n^*(\omega), P) < \alpha \quad (52)$$

compte tenu de (43), on en tire aisément :

$$\sup_{P \in \mathcal{D}} P^\infty \left[\bigcup_{n \geq N} d(P_n^*, P) \geq \alpha \right] \leq \frac{C(\alpha)}{N} \quad (53)$$

$\forall \alpha > 0$ et pour $N \geq n_0$ entier ne dépendant que de $\alpha^{(1)}$.

(53) entraîne évidemment la convergence presque sûre de P_n^* vers P , uniformément sur $\mathcal{D}$. CQFD.

Le théorème 5 peut se généraliser de la façon suivante :

COROLLAIRE 3 :

Si $\mathcal{P}$ est muni d'une topologie métrisable et séparable, si Q_n est un estimateur de P , mesurable et qui converge presque sûrement, uniformément sur $\mathcal{P}$, alors, il existe un estimateur de P , strict, mesurable, et qui converge presque sûrement, uniformément sur $\mathcal{D}$.

DÉMONSTRATION.

Analogue à celle du théorème 5.

Remarques.

1. Le corollaire 3 reste valable pour un paramètre à valeurs dans un espace métrique séparable quelconque ; il reste encore valable dans un espace métrique non séparable à condition de supposer que $Q_n(\Omega^n)$ est séparable (démonstrations analogues)⁽²⁾.

(1) Dans (53), $C(\alpha)$ est la constante associée, dans (43), à un voisinage fondamental de l'origine contenu dans la boule de centre 0 et de rayon $\alpha/3$.

(2) JILOVEC [10] p. 436 a énoncé un résultat très voisin.

2. Tous les résultats de ce paragraphe restent valables pour Ω non localement compact s'il est possible de munir $\mathcal{P}$ d'une topologie métrisable telle que :

- a) l'application $\omega \rightarrow \delta_{(\omega)}$ est mesurable ;
- b) $P_n(\Omega^n)$ est séparable ;
- c) $\forall P \in \mathcal{D}$, P_n tend vers P presque sûrement et cette convergence est uniforme sur $\mathcal{D}$.

Exemple : Les conditions précédentes sont vérifiées si Ω étant un espace Polonais on munit $\mathcal{P}$ de la topologie faible.

2. Estimation de $\varphi(P)$.

a) Préliminaires.

Commençons par indiquer deux lemmes très simples qui sont souvent utiles dans les problèmes d'estimation fonctionnelle.

LEMME 4 :

Soit (u_{nk}) une suite double de nombres réels telle que :

$$\lim_{n \rightarrow \infty} u_{nk} = 0 \quad \forall k \in \mathbb{N}.$$

Alors il existe une suite d'entiers $k(n)$ telle que

$$\lim_{n \rightarrow \infty} k(n) = +\infty \quad \text{et} \quad \lim_{n \rightarrow \infty} u_{nk(n)} = 0$$

DÉMONSTRATION.

Elémentaire.

LEMME 5 :

Si Θ est un espace topologique quelconque et si φ est continue

$\varphi(P_n^*)^{(1)}$ est un estimateur de $\varphi(P)$ strict, mesurable et qui converge presque sûrement, uniformément, sur $\mathcal{D}$.

DÉMONSTRATION :

Il est clair que $\varphi(P_n^*)$ est strict et mesurable. La deuxième partie est une conséquence de l'implication

$$\bigcap_{n \geq N} \{ P_n^* \in \varphi^{-1}(U) \} \implies \bigcap_{n \geq N} \{ \varphi(P_n) \in U \}$$

$\forall U$ voisinage ouvert de $\varphi(P)$, $\forall P \in \mathcal{D}$; et de (53) .

Donnons quelques exemples d'application du lemme 5 .

Exemples :

1. $\Omega = \mathbb{R}^P$, $\mathcal{D} = \mathcal{P}$, $\Theta = C(K, \mathbb{C})$ (K compact de $\mathbb{R}^P$) muni de la topologie de la convergence uniforme.

$\varphi(P)$ = restriction à K de la fonction caractéristique de P . On voit, d'après le théorème de Paul Lévy, que le lemme 5 s'applique :

$$\sup_{t \in K} \left| \frac{1}{n} \sum_{j=1}^n e^{i \langle t, X_j \rangle} - \varphi(P)(t) \right| \xrightarrow[p.s.]{} 0 \quad (54)$$

uniformément sur $\mathcal{P}^{(2)}$.

2. $\Omega = \mathbb{R}$, $\mathcal{D} = \mathcal{P}$, Θ = espace des fonctions de répartition à une variable, muni de la distance δ de Paul Lévy. $\varphi(P) = F$ fonction de répartition de P . Alors, F_n désignant la fonction de répartition empirique :

$$\delta(F_n, F) \xrightarrow[p.s.]{} 0 \text{ uniformément sur } \mathcal{P}^{(3)} \quad (55)$$

(1) ou $\varphi(P_n)$ si $\mathcal{D}$ contient les mesures de probabilités discrètes.

(2) $(X_n)_{n \in \mathbb{N}}$ désigne une suite de variables aléatoires à valeurs dans $\mathbb{R}^P$, indépendantes et de même loi P .

(3) C'est aussi une conséquence du lemme 3 .

3. Exemple h du §III. En prenant $\mathcal{D} = \mathcal{P}_c$ (espace des lois de probabilité sur $\mathbf{R}$ à f.d.r. continue), on voit, à partir d'un théorème de Polya⁽¹⁾, que φ est continue : le lemme 5 s'applique à la *fonction de concentration*.

b) Construction d'un estimateur.

Nous supposons maintenant que $\varphi(P)$ prend ses valeurs dans un espace métrique et nous faisons sur φ l'hypothèse suivante : φ est une limite *simple* d'une suite (φ_k) d'applications uniformément continues de $\mathcal{D}$ dans Θ .

Soit $(k(n))$ une suite d'entiers telle que $\lim_{n \rightarrow \infty} k(n) = +\infty$; on définira une suite (T_n) d'estimateurs de $\varphi(P)$

en posant :

$$T_n = \varphi_{k(n)}(P_n^*), \quad \forall n \in \mathbf{N} \quad (56)$$

Nous obtenons un estimateur très général qui est couramment utilisé dans les Problèmes d'estimation fonctionnelle.

REMARQUE :

Dans la pratique, on effectuera souvent un « lissage » préalable de P_n^* pour améliorer la précision de l'estimation.

c) Convergence de T_n .

THEOREME 6 :

Si Θ est un espace métrique et si φ est limite simple d'une suite (φ_k) d'applications uniformément continues de $\mathcal{D}$ dans Θ ; alors, T_n est mesurable, et l'on peut choisir la suite $k(n)$ de telle sorte que T_n converge en probabilité vers $\varphi(P)$; de plus, si la convergence de (φ_k) est uniforme sur $\mathcal{D}'C\mathcal{D}$, il en est de même pour la convergence de T_n .

(1) cf, DUGUE [9] p, 32.

DÉMONSTRATION.

Tout d'abord, T_n est évidemment mesurable comme composée de deux applications mesurables.

Montrons maintenant la convergence :

Comme φ_k est uniformément continue, on peut écrire :

$$\begin{aligned} \forall \varepsilon > 0, \quad \alpha(k, \varepsilon) > 0 \quad \text{tel que} \\ d(P, Q) < \alpha(k, \varepsilon) \Rightarrow d'[\varphi_k(P), \varphi_k(Q)] < \varepsilon \quad \forall P, Q \in \mathcal{D} \\ \forall k \in \mathbb{N} \end{aligned} \quad (57)$$

où d' est la distance définie sur Θ .

De (57), on tire :

$$d(P_n^*, P) < \alpha(k, \frac{1}{k}) \Rightarrow d'[\varphi_k(P_n^*), \varphi_k(P)] < \frac{1}{k} \quad (58)$$

$$\forall P \in \mathcal{D}, \quad \forall \omega \in \Omega^n, \quad \forall k \in \mathbb{N}$$

d'où, en tenant compte de (53) :

$$\sup_{P \in \mathcal{D}} \text{Prob} \left[d'(\varphi_k(P_n^*), \varphi_k(P)) \geq \frac{1}{k} \right] \leq \frac{C'(k)}{n} \quad (59)$$

$$\text{où l'on a posé } C'(k) = C \left[\alpha(k, \frac{1}{k}) \right]$$

en appliquant le lemme 4 à la suite, $\frac{C'(k)}{n}$, on en déduit qu'il existe une suite d'entiers $k(n)$ telle que

$$\lim_{n \rightarrow \infty} k(n) = +\infty \quad \text{et}$$

$$\lim_{n \rightarrow \infty} \sup_{P \in \mathcal{D}} \text{Prob} \left[d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) \geq \frac{1}{k(n)} \right] = 0 \quad (60)$$

L'inégalité

$$d'(\varphi_{k(n)}(P_n^*), \varphi(P)) \leq d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) + d'(\varphi_{k(n)}(P), \varphi(P)) \quad (61)$$

et le fait que $\varphi_{k(n)}$ tend vers φ permettent de conclure.

La dernière partie du théorème est une conséquence immédiate de (60) et (61).

REMARQUE :

Si Θ est séparable ou si $\varphi_k(\mathcal{D})$ est séparable, $\forall K \in \mathbb{N}$ on en déduit, à partir d'un résultat de JILOVEC ([10] p. 366) *l'existence d'un estimateur mesurable presque sûrement convergent*. De plus, compte tenu de la remarque du §1, c), on peut imposer à cet estimateur d'être strict.

EXEMPLE :

$\Omega = \mathbb{R}$, $\mathcal{D}$ = espace des mesures de probabilités sur $\mathbb{R}$ absolument continues et à densité uniformément continue, $\Theta = C(\mathbb{R})$ muni de la topologie de la convergence uniforme. $\varphi(P) = f$ densité de P

Soit F la f.d.r. de P et soit (h_k) une suite qui tend vers 0 avec $\frac{1}{k}$, posons :

$$f_k(x) = \frac{F(x+h_k) - F(x)}{h_k} \quad \forall x \in \mathbb{R}, \quad \forall k \in \mathbb{N}$$

On définit ainsi une suite φ_k d'applications de $\mathcal{D}$ dans $C(\mathbb{R})$; ces applications sont uniformément continues et φ_k tend vers φ dans $C(\mathbb{R})$ quand k tend vers l'infini. Donc, le théorème 6 s'applique et l'on en déduit qu'il existe $(k(n))$ telle que $\lim_{n \rightarrow \infty} k(n) = +\infty$

et

$$\sup_{x \in \mathbf{R}} \left| f(x) - \frac{F_n^*(x + h_{k(n)}) - F_n^*(x)}{h_{k(n)}} \right| \xrightarrow{P} 0 \quad (62)$$

F_n^* désignant la f.d.r. de P_n^* .

DISCUSSION.

LE CAM-LORRAINE SCHWARTZ [11] et JILOVEC [10] ont obtenu, pour Ω quelconque, des théorèmes d'existence analogues au théorème 6. Les premiers ont traité le cas d'un espace métrique séparable, le second a traité le cas d'un espace de Hilbert.

Pour Ω localement compact, la structure uniforme définie dans ces deux articles et la structure uniforme associée à la topologie vague ne sont pas comparables en général.

La méthode que nous employons ici a l'avantage d'être simple et de fournir, à la fois, une condition d'existence et une méthode de construction d'un estimateur convergent et mesurable.

3. Cas particulier important : $\Omega = \mathbf{R}$.

Nous allons voir que, pour $\Omega = \mathbf{R}$, on peut améliorer le théorème 6.

Munissons $\mathcal{P}(\mathbf{R})$ de la distance de la convergence uniforme sur les fonctions de répartition. Pour des raisons de commodité, nous noterons encore d cette distance ⁽¹⁾.

Il est bien connu que P_n est alors une application mesurable ⁽²⁾ de $\mathbf{R}^n$ dans $\mathcal{P}(\mathbf{R})$; de plus, $P_n(\mathbf{R}^n)$ est séparable ⁽³⁾.

Commençons par énoncer un lemme, analogue au lemme 3, mais plus précis :

(1) Bien qu'elle n'induisse pas la topologie vague sur $\mathcal{P}(\mathbf{R})$.

(2) Mais non continue.

(3) On peut recouvrir $P_n(\mathbf{R}^n)$ avec les boules ouvertes centrées sur les $P_n(\omega_1, \dots, \omega_n)$ correspondant aux points de coordonnées rationnelles, et de rayon $2/n$.

LEMME 6 :

(DVORETZKY - KIEFER - WOLFOWITZ - SCHUSTER) :

On a l'inégalité

$$P^\infty [d(P_n, P) > \varepsilon] \leq C e^{-2n\varepsilon^2}, \quad \forall P \in \mathcal{P} \quad \forall \varepsilon > 0 \quad (63)$$

$$\forall n \in \mathbb{N}$$

où C est une constante universelle.

DÉMONSTRATION.

Voir DVORETZKY - KIEFER - WOLFOWITZ [13] et SCHUSTER [14].

On a alors le résultat suivant :

THÉORÈME 7 :

Si $\Omega = \mathbb{R}$, si Θ est un espace métrique et si φ est limite simple d'une suite (φ_k) d'applications uniformément continues de $\mathcal{D}$ dans Θ , alors, T_n est mesurable, et l'on peut choisir la suite $k(n)$ de telle sorte que T_n converge presque complètement sûrement⁽¹⁾ vers $\varphi(P)$.

DÉMONSTRATION.

La construction de P_n^* est identique à celle du §1 b) ; l'implication (52) reste valable, ce qui permet d'écrire, en tenant compte de (63) :

$$\sup_{P \in \mathcal{D}} \text{Prob} [d(P_n^*, P) > \alpha] \leq C e^{-\frac{2}{3} n \alpha^2}, \quad \forall n \in \mathbb{N}, \quad \forall \alpha > 0 \quad (64)$$

alors, (58) et (64) entraînent

(1) Donc presque sûrement.

$$\sup_{P \in \mathcal{T}} \text{Prob} \left[d'(\varphi_k(P_n^*), \varphi_k(P)) > \frac{1}{k} \right] \leq C e^{-\frac{2}{3} n \alpha(k, \frac{1}{k})^2} \quad (65)$$

$$\forall k, n \in \mathbb{N}.$$

D'après le lemme 4, on peut choisir $(k(n))$ telle que $\lim_{n \rightarrow \infty} k(n) = +\infty$ et

$$\lim_{n \rightarrow \infty} \frac{1}{[\sqrt{n} \alpha(k(n), \frac{1}{k(n)})]^2} = 0$$

On aura donc, pour n assez grand :

$$\sup_{P \in \mathcal{T}} \text{Prob} \left[d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) > \frac{1}{k(n)} \right] \leq C e^{-\frac{2}{3} \cdot \frac{\sqrt{n}}{\delta}} \quad (66)$$

($\delta > 0$ donné), on en déduit que

$$\sum_{n=1}^{\infty} \text{Prob} [d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) > \varepsilon] < +\infty \quad (67)$$

$\forall P \in \mathcal{T}, \forall \varepsilon > 0$; d'où le résultat compte tenu de (61).

EXEMPLE :

Dans (62) on peut remplacer la convergence en probabilité par la convergence presque complètement sûre, $k(n)$ étant convenablement choisi.

V. ESTIMATION SYMÉTRIQUE.

Nous terminons ce chapitre en généralisant les résultats de HALMOS [5] sur les estimateurs symétriques sans biais.

1. Unicité.

Soit T_n un estimateur d'ordre n de $\varphi(P)$; on sait que l'estimateur *symétrisé* de T_n est défini par :

$$T_n^{[n]} = S_n = \frac{1}{N} \sum T_n(\omega_{i_1}, \dots, \omega_{i_n}) \quad \forall (\omega_1, \dots, \omega_n) \in \Omega^n \quad (68)$$

où la somme porte sur le nombre de permutations distinctes des ω_i et où N est le nombre de ces permutations.

Il est clair que, si T_n est sans biais, S_n est également sans biais : on peut donc toujours, quand on connaît un estimateur sans biais de $\varphi(P)$ construire un estimateur *symétrique* sans biais de $\varphi(P)$. Nous allons voir que, sous des conditions assez générales, cet estimateur symétrique sans biais est unique.

Donnons, tout d'abord, deux définitions et un résultat préliminaire qui généralise des résultats de HALMOS et de FRASER :

DÉFINITION 11:

On dira que l'ensemble $\mathcal{D}$ est de type F s'il contient les mesures de probabilité à support fini.

DÉFINITION 11 bis :

Pour $\Omega = \mathbf{R}$, on dira que l'ensemble $\mathcal{D}$ est de type C s'il contient les lois absolument continues à densité uniforme sur chaque intervalle d'une famille finie d'intervalles disjoints.

LEMME 7 :

Ω étant localement compact et Θ étant un E. V. T. localement convexe séparé soit $t(\omega_1, \dots, \omega_n)$ une application symétrique, P^n -intégrable $\forall P \in \mathcal{D}$ de Ω^n dans Θ , alors :

1° - Si $\mathcal{D}$ est de type F :

$$E_p [t(\omega_1, \dots, \omega_n)] = 0, \quad \forall P \in \mathcal{D} \Rightarrow t(\omega_1, \dots, \omega_n) = 0 \\ \forall (\omega_1, \dots, \omega_n) \in \Omega^n. \quad (69)$$

2) - Si $\mathcal{D}$ est de type C, et si t est continue :

$$E_p [t(\omega_1, \dots, \omega_n)] = 0, \quad \forall P \in \mathcal{D} \Rightarrow t(\omega_1, \dots, \omega_n) = 0 \quad (70) \\ \forall (\omega_1, \dots, \omega_n) \in \mathbb{R}^n.$$

DÉMONSTRATION.

Remarquons que le lemme 1 de [5] reste valable pour un polynôme à coefficients dans un espace vectoriel. La démonstration de la première partie du lemme 7 est donc tout à fait analogue à celle du lemme 2 de [5].

Pour démontrer la deuxième partie, nous procéderons de la façon suivante : Soit $\omega^0 = (\omega_1^0, \dots, \omega_n^0)$ un point de $\mathbb{R}^n$ à coordonnées deux à deux distinctes et soit $h > 0$ assez petit pour que les intervalles

$$I_j = [\omega_j^0 - h, \omega_j^0 + h], \quad j = 1, \dots, n$$

soient deux à deux disjoints. Soient $p_1, \dots, p_n$ n nombres positifs tels que :

$$\sum_{j=1}^n p_j = 1;$$

Enfin, soit P_h la mesure de probabilité à densité uniforme et égale à $\frac{p_j}{2h}$ sur I_j , $j = 1, \dots, n$. Alors :

$$E_{p_j} [t(\omega_1, \dots, \omega_n)] = \sum_{j_1, \dots, j_n} \int_{I_{j_1} \times \dots \times I_{j_n}} t(\omega_1, \dots, \omega_n) \frac{p_{j_1} \dots p_{j_n}}{(2h)^n} d\omega_1 \dots d\omega_n$$

est un polynôme homogène par rapport aux variables $p_1, \dots, p_n$; d'après l'hypothèse faite sur t et le lemme 1 de [5] il est identiquement nul. D'autre part, comme t est symétrique, le coefficient de $p_1 \dots p_n$ s'écrit :

$$\frac{n!}{(2h)^n} \int_{I_1 \times \dots \times I_n} t(\omega_1, \dots, \omega_n) d\omega_1, \dots, d\omega_n$$

donc :

$$\int_{I_1 \times \dots \times I_n} t(\omega_1, \dots, \omega_n) d\omega_1 \dots d\omega_n = \vec{0}^{(1)} \quad (71)$$

Ce qui, d'après (12) implique :

$$\int_{I_1 \times \dots \times I_n} \langle t(\omega_1, \dots, \omega_n), \theta' \rangle dP_h^n = 0, \quad \forall \theta' \in \Theta' \quad (72)$$

Ce qui s'écrit encore :

$$\int_{I_1 \times \dots \times I_n} (\theta' \circ t)(\omega_1, \dots, \omega_n) dP_h^n(\omega_1, \dots, \omega_n) = 0 \quad \forall \theta' \in \Theta' \quad (73)$$

t étant continue $\theta' \circ t$ l'est également ; donc, il existe ω^h (dépendant de θ'), $\in I_1 \times \dots \times I_n$ et tel que $(\theta' \circ t)(\omega^h) = 0$. Le raisonnement précédent étant valable $\forall h > 0$ assez petit, on peut donc construire une suite $(\omega^{h(k)})_{k \in \mathbb{N}}$ qui tend vers ω^0 et telle que $(\theta' \circ t)(\omega^{h(k)}) = 0, \forall k \in \mathbb{N}$, on en déduit :

$$(\theta' \circ t)(\omega^0) = 0 \quad \forall \theta' \in \Theta'$$

d'où l'on tire ⁽²⁾

$$t(\omega^0) = \vec{0}.$$

(1) $\vec{0}$ désignant ici l'élément neutre de Θ .

(2) C'est une conséquence du théorème de Hahn-Banach, cf, BOURBAKI, [15] Chapitre IV, p. 48.

Donc t s'annule en tout point de $\mathbf{R}^n$ à coordonnées deux à deux distinctes et comme t est continue, il vient :

$$t(\omega) \equiv \vec{0}, \quad \text{CQFD}^{(1)}.$$

Nous pouvons donner maintenant une généralisation du théorème 3 de [5].

THÉORÈME 8 :

Si Θ est un E.V.T. localement convexe séparé, si $\mathcal{D}$ est de type F (resp, de type C) et si T_n est un estimateur de $\varphi(P)$ sans biais (resp, continu et sans biais) alors, le seul estimateur de $\varphi(P)$ symétrique et sans biais (resp, symétrique, continue et sans biais) est $T_n^{[n]}$.

DÉMONSTRATION.

S_n désignant un estimateur symétrique, sans biais (resp, symétrique, continu et sans biais) de $\varphi(P)$, on voit que $S_n - T_n^{[n]}$ vérifie les conditions du lemme 7, 1°) (resp, du lemme 7, 2°), donc $S_n = T_n^{[n]}$,

CQFD.

2. Optimalité.

Nous allons voir, maintenant, qu'en symétrisant un estimateur, on diminue le risque associé (la fonction de risque étant définie par la relation (6)), Le théorème ci-dessous est une généralisation du théorème 5 de [5].

THÉORÈME 9 :

Sous les conditions du théorème 8 et si Θ est un espace de Hilbert réel, l'estimateur symétrique sans biais (resp. symétrique continu et sans biais) est uniformément optimal dans la famille des estimateurs sans

(1) Pour $\Theta = \mathbf{R}$, on peut s'affranchir de la condition de continuité. Cf, FRASER [16] p, 28-30,

sans biais (resp, continu et sans biais) pour la fonction de risque $R(T_n, \varphi(P)) = E_P ||T_n - \varphi(P)||^2$.

DÉMONSTRATION.

Si Θ est un Hilbert réel, on peut écrire :

$$\begin{aligned} E_P ||T_n - E_P(T_n)||^2 &= E_P \langle T_n - E_P(T_n), T_n - E_P(T_n) \rangle \\ &= E_P ||T_n||^2 - ||E_P(T_n)||^2 \end{aligned} \quad (74)$$

Comparer les risques associés à deux estimateurs sans biais revient donc à comparer les moments d'ordre 2 de leurs normes.

Or, si T_n désigne un estimateur sans biais (resp, continu et sans biais) de $\varphi(P)$ et si S_n désigne l'estimateur symétrique sans biais (resp, symétrique continu et sans biais) de $\varphi(P)$, on peut écrire, d'après (78) :

$$S_n(\omega_1, \dots, \omega_n) = \frac{1}{N} \sum T_n(\omega_{i_1}, \dots, \omega_{i_n}) = \frac{1}{N} \sum_{i=1}^N T_{ni} \quad (75)$$

$$\text{d'où } ||S_n|| \leq \frac{1}{N} \sum_{i=1}^N ||T_{ni}|| \quad (76)$$

$$\text{soit } ||S_n||^2 \leq \left(\frac{1}{N} \sum_{i=1}^N ||T_{ni}|| \right)^2 \leq \frac{1}{N} \sum_{i=1}^N ||T_{ni}||^2 \quad (77)$$

$$\text{d'où } E_P ||S_n||^2 \leq \frac{1}{N} \sum_{i=1}^N E_P ||T_{ni}||^2 = E_P ||T_n||^2 \quad (78)$$

C. Q. F. D.

Exemple :

$\Omega = [0, 1]$; $\Theta = L^2[0, 1]$, la f.d.r. empirique F_n est un estimateur uniformément optimal dans la famille des estimateurs sans biais de la f.d.r. Un calcul facile montre que :

$$R(F_n, F) = E \int_0^1 [F_n(x) - F(x)]^2 dx = \frac{1}{n} \int_0^1 F(x) [1 - F(x)] dx \quad (79)$$

COROLLAIRE 4 :

Si Θ est un espace de Hilbert réel et si T_n est un estimateur de $\varphi(P)$, P^n -intégrable, $\forall P \in \mathcal{D}$, alors $T_n^{[n]}$ est P^n -intégrable, $\forall P \in \mathcal{D}$ et

$$R[T_n^{[n]}, \varphi(P)] \leq R[T_n, \varphi(P)], \quad \forall P \in \mathcal{D} \quad (80)$$

DÉMONSTRATION :

Analogue à celle du théorème 9.

REMARQUES :

1. Le corollaire 4 justifie le choix d'estimateurs symétriques par rapport aux observations ; il est clair que ce choix est naturel puisque le numérotage des observations est complètement arbitraire. Il semble intéressant de poser le problème de l'estimation d'un paramètre (à partir d'un échantillon de taille n) d'une manière qui ne fasse pas intervenir le numérotage des observations : on peut, par exemple, utiliser la notion de répartition ponctuelle aléatoire⁽¹⁾. Nous comptons revenir sur cette question.

2. On pourrait retrouver le théorème 9 à partir d'une généralisation des théorèmes de RAO-BLACKWELL et LEHMANN-SCHEFFE. L'examen des démonstrations de ces théorèmes (cf. FRASER [16] p. 57-58 et p. 61) montre en effet qu'on pourrait les généraliser au cas où Θ est un E.V.T. localement convexe séparé.

(1) cf. LANNUZEL [17].

II

ESTIMATION DE LA DENSITÉ DE PROBABILITÉ.

1 - GÉNÉRALITÉS.

1. La méthode du noyau.

Le problème de l'estimation de la densité de probabilité a été l'objet de nombreux travaux au cours des dernières années.

Le premier article important sur cette question semble être celui de ROSENBLATT [7] (1956) ; dans cet article, l'auteur propose une méthode d'estimation qui a été très étudiée par la suite : *la méthode du noyau*. Cette méthode consiste à prendre comme estimateur de la densité la fonction aléatoire $\hat{f}_n$ définie de la façon suivante :

$$\hat{f}_n(\omega; \omega_1, \dots, \omega_n) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{\omega - \omega_i}{h}\right), \quad \forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (1)$$

où le noyau K est souvent une densité possédant certaines propriétés de régularité et h un nombre positif qui dépend de n .

Cet estimateur a été étudié notamment par PARZEN [29] (1962), NADARAYA [18] (1965) et WOODROOFE [19] (1967) pour $\Omega = \mathbf{R}$; par CACOULOS [20] (1966) et VAN RYZIN [21] (1969) pour $\Omega = \mathbf{R}^p$ et par CRASWELL [22] (1965) dans un groupe topologique.

LEADBETTER et WATSON [23] (1963) ont proposé une méthode un peu plus générale ; ils considèrent un estimateur de la forme :

$$\hat{f}_n(\omega; \omega_1, \dots, \omega_n) = \frac{1}{n} \sum_{i=1}^n K_n(\omega - \omega_i), \quad \forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (2)$$

leurs résultats ont été complétés récemment par PICKANDS III [24] (1969).

La Méthode du noyau a certaines propriétés intéressantes (notamment la rapidité de convergence et le fait que l'estimateur est strict), mais

elle a, pour un praticien, certains inconvénients (complication et manque de robustesse des estimateurs optimaux).

2. La méthode des fonctions orthogonales.

La méthode des fonctions orthogonales, que nous étudierons dans ce chapitre, a été introduite par CENCOV [25] (1962)⁽¹⁾. Le principe de cette méthode est le suivant :

Soit μ une mesure sur $(\Omega, \mathcal{F})$ et soit $\mathcal{D}$ l'ensemble des mesures de probabilité sur $(\Omega, \mathcal{F})$ admettant par rapport à μ , une densité de Radon-Nikodym appartenant à $L^2(\mu)$.

On sait que, muni de la norme $\int_{\Omega} |f|^2 d\mu^{1/2}$

$L^2(\mu)$ est un espace de Hilbert ; soit alors $(e_i)_{i \in I}$

une base orthonormale de $L^2(\mu)$ et soit $(e_{i_k, n})_{1 \leq k \leq q(n)}$

un sous-ensemble fini de (e_i) ; une version $\hat{f}_n^*$ de l'estimateur $\hat{f}_n$ de $f = \frac{dP}{d\mu}$ sera défini par la relation :

$$\hat{f}_n^*(\omega; \omega_1, \dots, \omega_n) = \sum_{k=1}^{q(n)} \frac{1}{n} \left[e_{i_k, n}^0(\omega_1) + \dots + e_{i_k, n}^0(\omega_n) \right] e_{i_k, n}^*(\omega),$$

$$\forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (3)$$

où les $e_{i_k, n}^*$ ($k = 1, \dots, q(n)$) sont des versions des

$e_{i_k, n}$ ($k = 1, \dots, q(n)$) et les $e_{i_k, n}^0$ des versions fixées des $e_{i_k, n}$.

(1) Et, indépendamment, par TIAGO DE OLIVEIRA [26].

On voit que cette méthode consiste à estimer la projection de $\frac{dP}{d\mu}$ sur un sous-espace de $L^2(\mu)$ de dimension finie : elle ramène l'estimation de $\frac{dP}{d\mu}$ à celle d'un nombre fini de paramètres numériques.

Remarquons, d'autre part, que le principe de cette méthode peut s'appliquer à l'estimation de n'importe quel paramètre à valeurs dans un espace de Hilbert.

Depuis 1962 l'estimateur défini par (3) a été peu étudié ; cependant, deux cas particuliers importants ont été abordés : S. C. SCHWARTZ [27] (1967) a pris, comme système orthonormal dans L^2 , les fonctions d'Hermite alors que KRONMAL et TARTER [28] (1968) s'intéressaient au cas des fonctions trigonométriques.

Dans la suite, nous généralisons leurs résultats et nous donnons, dans le cas général, des résultats qui, à notre connaissance, sont nouveaux (en particulier ceux de notre note aux CR du 10 mars 1969 [32]).

II - CONVERGENCE EN MOYENNE QUADRATIQUE INTÉGRÉE.

Les conditions de convergence du MISE (défini au § II du chapitre I) sont simples et connues⁽¹⁾, nous les énoncerons sous la forme suivante:

THÉORÈME 10 :

Si les séquences $(e_{i_{k,n}})_{1 \leq k \leq q(n)}$ sont telles que la suite

$$\sum_{k=1}^{q(n)} \left[\int_{\Omega} e_{i_{k,n}}(\omega) f(\omega) d\mu(\omega) \right]^2 \quad \text{converge vers}$$

(1) Car on peut considérer qu'elles sont contenues, implicitement, dans les théorèmes 1 a et 1 b de Cencov [25] .

$\int_{\Omega} f(\omega)^2 d\mu(\omega)^{(1)}$, une condition nécessaire et suffisante pour que $\hat{f}_n$ converge en moyenne quadratique intégrée vers f est que

$$\lim_{n \rightarrow \infty} \int_{\Omega} \frac{1}{n} \left[\sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) d\mu(\omega) = 0 \quad (4)$$

DÉMONSTRATION

Pour simplifier les notations, nous poserons

$$a_i = \int_{\Omega} e_i(\omega) f(\omega) d\mu(\omega) \quad \forall i \in I$$

$$a_{i_{k,n}} = \frac{1}{n} \sum_{j=1}^n e_{i_{k,n}}^0(\omega_j) \quad \forall k = 1, \dots, q(n); \quad \forall n \in \mathbb{N}.$$

et

$$\sum_{k=1}^{q(n)} a_{i_{k,n}} \cdot e_{i_{k,n}} = f_n \quad \forall n \in \mathbb{N}$$

alors puisque $\hat{f}_n - f_n$ et $f_n - f$ sont orthogonaux, le MISE peut s'écrire :

$$J_n = E \int_{\Omega} [\hat{f}_n(\omega) - f_n(\omega)]^2 d\mu(\omega) + \int_{\Omega} [f_n(\omega) - f(\omega)]^2 d\mu(\omega) \quad (5)$$

soit

$$J_n = \frac{1}{n} \sum_{k=1}^{q(n)} \left[\int_{\Omega} e_{i_{k,n}}^2(\omega) f(\omega) d\mu(\omega) - a_{i_{k,n}}^2 \right] + \sum_{j \in I_n} a_j^2 \quad (6)$$

où l'on a posé $I_n = I - \{i_{1,n}, \dots, i_{q(n),n}\}$

ou encore :

(1) Dans la suite, nous appellerons cette condition "condition S".

$$J_n = \int_{\Omega} \frac{1}{n} \left[\sum_{k=1}^{q(n)} e_{i_k, n}^2(\omega) \right] f(\omega) d\mu(\omega) - \frac{n+1}{n} \sum_{k=1}^{q(n)} a_{i_k, n}^2 + \sum_{j \in I} a_j^2 \quad (7)$$

Or l'hypothèse faite sur les séquences $(e_{i_k, n})$ implique que

$$\lim_{n \rightarrow \infty} \frac{n+1}{n} \sum_{k=1}^{q(n)} a_{i_k, n}^2 = \sum_{j \in I} a_j^2 \quad (8)$$

donc, d'après (7) et (8), $J_n \rightarrow 0$ si et seulement si on a (4). CQFD.

La condition (4) dépend étroitement de f , cependant nous allons voir que, dans de nombreux cas particuliers importants, cette dépendance n'est qu'apparente :

COROLLAIRE 5 :

1°) - Si l'une des conditions suivantes est remplie :

- a) Il existe une version des e_i uniformément bornée pour $i \in I$ et $\omega \in \Omega$.
- b) Il existe $g \in L^4(\mu)$ qui majore $|e_i|$ pour tout $i \in I$.
- c) Il existe une version de f uniformément bornée pour $\omega \in \Omega$.

Alors, si la condition S est vérifiée,

$$\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0 \quad \text{entraîne} \quad \lim_{n \rightarrow \infty} J_n = 0.$$

2°) - S'il existe une version de f uniformément minorée par un nombre positif m pour $\omega \in \Omega$ et si la condition S est vérifiée, alors :

$$\lim_{n \rightarrow \infty} J_n = 0 \quad \text{entraîne} \quad \lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0.$$

DÉMONSTRATION.

Les résultats proviennent des implications suivantes :

$$1^{\circ}) \text{ a) } \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) = O(1)$$

$$\Rightarrow \int_{\Omega} \frac{1}{n} \left[\sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) d\mu(\omega) = O\left(\frac{q(n)}{n}\right)$$

$$\begin{aligned} \text{b) } \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) &\leq \int_{\Omega} g^2(\omega) f(\omega) d\mu(\omega) \\ &\leq \left[\int_{\Omega} g^4(\omega) d\mu(\omega) \right]^{1/2} \cdot \left[\int_{\Omega} f^2(\omega) d\mu(\omega) \right]^{1/2} \end{aligned}$$

$$\text{c) } \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) = O(1)$$

$$2^{\circ}) \inf_{\omega \in \Omega} f^*(\omega) \geq m \Rightarrow \int_{\Omega} \frac{1}{n} \left[\sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) d\mu(\omega) \geq m \cdot \frac{q(n)}{n}$$

COROLLAIRE 6 :

On suppose que Ω est un intervalle compact de $\mathbf{R}$, que μ est la mesure de Lebesgue sur Ω et que les e_i ($i \in \mathbf{N}$) sont les fonctions trigonométriques sur Ω , alors, si

$$e_{i_{k,n}} = e_k, \quad \forall k \in \mathbf{N}, \quad \forall n \in \mathbf{N}$$

et si $\lim_{n \rightarrow \infty} q(n) = +\infty$; pour que $\lim_{n \rightarrow \infty} J_n = 0$,

il faut et il suffit que $\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0$.

DÉMONSTRATION.

On peut, sans perte de généralité, supposer que $\Omega = [0, 2\pi]$; on aura alors :

$$e_1^*(\omega) = \frac{1}{\sqrt{2\pi}}, \quad e_2^*(\omega) = \frac{\cos \omega}{\sqrt{\pi}}, \quad e_3^*(\omega) = \frac{\sin \omega}{\sqrt{\pi}},$$

$$e_4^*(\omega) = \frac{\cos 2\omega}{\sqrt{\pi}} \quad \text{etc}$$

Un calcul élémentaire montre que :

$$\frac{q(n) - 1}{2\pi} \leq \sum_{k=1}^{q(n)} e_k^{*2}(\omega) \leq \frac{q(n) + 1}{2\pi} \left[\forall q(n) \in \mathbb{N}, \forall \omega \in \Omega \right] \quad (9)$$

Or, comme $\lim_{n \rightarrow \infty} q(n) = +\infty$, la condition S est vérifiée, donc (4) est une condition nécessaire et suffisante de convergence, et comme, d'après (9), (4) est équivalent à $\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0$, le corollaire est démontré.

REMARQUE :

Le corollaire 6 améliore la deuxième partie du théorème 3 de [28] où l'on montre que $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} = 0$ est une condition suffisante pour que

$$\lim_{n \rightarrow \infty} J_n = 0.$$

Optimalité et vitesse de convergence.

La relation (7) montre que J_n dépend fortement des coefficients a_i ; il n'est donc pas possible, en général, d'obtenir un estimateur uniformément optimal (pour la fonction de risque J_n) dans la famille des estimateurs définis par (3); c'est ce que confirme le

LEMME 8 :

On suppose que la famille e_i est dénombrable, que $e_{i_k, n} = e_k$, $\forall k \in \mathbb{N}$,

$\forall n \in \mathbb{N}$ et qu'il existe deux constantes m et M telles que

$$0 < m \leq \int_{\Omega} e_i^{*2}(\omega) f(\omega) d\mu(\omega) \leq M, \quad \forall i \in \mathbb{N} \quad (10)$$

Dans ces conditions et $\forall n \in \mathbb{N}$, il existe un entier $\bar{q}_0(n)$ qui minimise J_n ; de plus, si $|a_i|$ est fonction décroissante de i , $\bar{q}_0(n)$ est tel que,

au moins pour n assez grand :

$$|a_{q_0(n)}| \geq \frac{\sqrt{m}}{\sqrt{n+1}} \text{ et } |a_{q_0(n)+1}| \leq \frac{\sqrt{m}}{\sqrt{n+1}} \quad (11)$$

DÉMONSTRATION.

Désignons par $\Delta_q J_n$ l'accroissement de J_n quand $q(n)$ passe de q à $q+1$; (7) entraîne alors :

$$\Delta_q J_n = \frac{1}{n} \left[\int_{\Omega} e^{2q+1}(\omega) f(\omega) d\mu(\omega) - (n+1) a_{q+1}^2 \right] \quad (12)$$

d'où, en tenant compte de (10) :

$$\frac{1}{n} [m-(n+1) a_{q+1}^2] \leq \Delta_q J_n \leq \frac{1}{n} [M-(n+1) a_{q+1}^2] \quad (13)$$

Donc, $\forall n \in \mathbb{N}$, il existe un entier $q_1(n)$ tel que

$$q \geq q_1(n) \implies |a_{q+1}| < \frac{\sqrt{m}}{\sqrt{n+1}} \implies \Delta_q J_n > 0 \quad (14)$$

c'est-à-dire que, pour q assez grand, J_n est fonction croissante de q : la première partie du lemme est démontrée. Remarquons que l'on a aussi montré que $q_0(n) \leq q_1(n)$.

Supposons maintenant que $|a_q|$ soit fonction décroissante de q , alors si $q_1(n)$ est le plus petit entier tel que l'on ait (14), on en déduit que $|a_{q_0(n)}| \geq \frac{\sqrt{m}}{\sqrt{n+1}}$, au moins pour n assez grand.

Soit $q_2(n)$ le plus grand entier tel que $|a_{q+1}| > \frac{\sqrt{m}}{\sqrt{n+1}}$ ($q_2(n)$ existe au moins pour n assez grand), d'après (13), on a l'implication :

$$q \leq q_2(n) \implies \Delta_q J_n < 0$$

On en déduit $q_0(n) > q_2(n)+1$; d'où, d'après l'hypothèse faite sur $q_2(n)$,

$$|a_{q_0(n)+1}| \leq \frac{\sqrt{M}}{\sqrt{n+1}}, \text{ ce qui achève la démonstration.}$$

Le lemme 8 montre qu'il faut connaître l'ordre de grandeur des coefficients a_q pour connaître celui de J_n .

On peut, cependant, obtenir des résultats partiels : par exemple si

$|a_k| = O\left(\frac{1}{k^r}\right)$, $r > \frac{1}{2}$, et si l'on prend $q(n) = O(n^{1/2r})$ on voit

facilement que $J_n = O\left(\frac{1}{n^{2r-1/2r}}\right)$; si, de plus, $|a_k|$ est de l'ordre

de $\frac{1}{k^r}$, on obtient ainsi un estimateur asymptotiquement optimal au sens suivant : J_n est asymptotiquement du même ordre que le MISE associé à l'estimateur optimal en f .

Remarquons enfin que l'on peut évaluer empiriquement $q_0(n)$ en remplaçant J_n par une estimation $\hat{J}_n$ et en étudiant le signe de $\Delta_q \hat{J}_n$. Cette méthode a été employée par KRONMAL et TARTER [28].

III - CONVERGENCE UNIFORME EN MOYENNE QUADRATIQUE ET CONVERGENCE UNIFORME PRESQUE SURE.

1. Introduction.

Dans les problèmes d'estimation de la densité, on cherche souvent à estimer une version spécifiée de la densité; il s'agira, en général, d'une version *continue* de la densité. On peut alors étudier des modes de convergence plus précis que celui du paragraphe précédent.

Rappelons d'abord à quelle condition une version continue de la densité est unique :

LEMME 9 :

Soit Ω un espace topologique muni de sa tribu Borélienne, soit μ une mesure sur Ω qui charge tout ouvert non vide et soit P une mesure de probabilité sur Ω admettant, par rapport à μ , une densité de Radon-Nikodym $\frac{dP}{d\mu}$. Alors, s'il existe une version continue de $\frac{dP}{d\mu}$, cette version est unique.

DÉMONSTRATION.

Soient f et g deux versions continues de $\frac{dP}{d\mu}$ et soit $\omega_0 \in \Omega$ tel que $f(\omega_0) - g(\omega_0) > 0$, alors, il existe un voisinage ouvert U de ω_0 tel que l'on ait simultanément :

$$\int_U \frac{dP}{d\mu}(\omega) d\mu(\omega) = \int_U f(\omega) d\mu(\omega) = \int_U g(\omega) d\mu(\omega)$$

$$\text{et } \int_U [f(\omega) - g(\omega)] d\mu(\omega) > 0$$

ce qui est évidemment contradictoire. CQFD.

EXEMPLE :

Les conditions du lemme 8 sont vérifiées si Ω est un groupe topologique localement compact et si μ est une mesure de Haar sur Ω .

REMARQUES :

1.- Si μ s'annule sur un ouvert non vide de Ω , s'il existe une version continue de $\frac{dP}{d\mu}$ et si Ω est normal, il existe, en général, une infinité de versions continues de $\frac{dP}{d\mu}$.

2.- Dans la suite, si les conditions du lemme 9 sont vérifiées, nous disons " P admet une densité continue " au lieu de " il existe une version continue de $\frac{dP}{d\mu}$ " ; de même, nous dirons " e_i est continue " au lieu de " il existe une version continue de e_i " ⁽¹⁾.

2. Convergence uniforme en moyenne quadratique.

THÉORÈME 11 :

Sous les conditions suivantes :

(1) Car le lemme 9 s'applique à e_i^2 donc à e_i .

1°) f est continue

2°) e_i est continue et bornée, $\forall i \in I$

3°) La famille $\sum_{i \in I} a_i e_i$ est uniformément sommable de somme f .

4°) Les séquences $(e_{i_k, n})_{1 \leq k \leq q(n)}$ sont telles que la suite

$$f_n = \sum_{k=1}^{q(n)} a_{i_k, n} e_{i_k, n} \text{ converge uniformément vers } f^{(1)}.$$

$$\text{Si } \lim_{n \rightarrow \infty} \left[\sup_{1 \leq k \leq q(n)} \sup_{\omega \in \Omega} |e_{i_k, n}(\omega)| \right]^4 \cdot \frac{q(n)^2}{n} = 0 \quad (15)$$

$$\text{Alors } Q_n = \sup_{\omega \in \Omega} E [|\hat{f}_n(\omega) - f(\omega)|^2] \rightarrow 0 \text{ quand } n \rightarrow +\infty \quad (16)$$

DÉMONSTRATION.

$$\text{Posons pour simplifier } \sup_{1 \leq k \leq q(n)} \sup_{\omega \in \Omega} |e_{i_k, n}(\omega)| = M_n.$$

D'après la relation

$$E [|\hat{f}_n(\omega) - f(\omega)|^2] = E [|\hat{f}_n(\omega) - f_n(\omega)|^2] + |f_n(\omega) - f(\omega)|^2$$

et la condition S' , il suffit de montrer que

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} E [|\hat{f}_n(\omega) - f_n(\omega)|^2] = 0.$$

$$\text{Or } \hat{f}_n(\omega) - f_n(\omega) = \sum_{k=1}^{q(n)} (\hat{a}_{i_k, n} - a_{i_k, n}) e_{i_k, n}(\omega)$$

$$\text{D'où } E [|\hat{f}_n(\omega) - f_n(\omega)|^2] = \sum_{k=1}^{q(n)} \text{Var}(\hat{a}_{i_k, n}) e_{i_k, n}^2(\omega)$$

$$+ \sum_{k \neq k'} \text{Cov}(\hat{a}_{i_k, n}; \hat{a}_{i_{k'}, n}) e_{i_k, n}(\omega) e_{i_{k'}, n}(\omega).$$

On en tire

$$E [|\hat{f}_n(\omega) - f_n(\omega)|^2] \leq M_n^2 \left[\sum_{k=1}^{q(n)} \text{Var}(\hat{a}_{i_k, n}) + \sum_{k \neq k'} |\text{Cov}(\hat{a}_{i_k, n}; \hat{a}_{i_{k'}, n})| \right]$$

(1) Dans la suite, nous appellerons cette condition "condition S' ".

d'où :

$$E [|\hat{f}_n(\omega) - f_n(\omega)|^2] \leq M_n^2 \left[\sum_{k=1}^{q(n)} \text{Var}(\hat{a}_{i_k, n}) + \sum_{k \neq k'} \sqrt{\text{Var}(\hat{a}_{i_k, n}) \cdot \text{Var}(\hat{a}_{i_{k'}, n})} \right]$$

soit, en posant $\sigma^2(e_i) = \text{Var } e_i$, $i \in I$:

$$E [|\hat{f}_n(\omega) - f_n(\omega)|^2] \leq M_n^2 \cdot \left[\sum_{k=1}^{q(n)} \frac{\sigma(e_{i_k, n})}{\sqrt{n}} \right]^2 \leq M_n^4 \cdot \frac{q(n)^2}{n}$$

et la condition (10) permet de conclure.

REMARQUES :

1.- Si les e_i sont bornés, uniformément pour $i \in I$, la condition (15) s'écrit simplement

$$\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} = 0 \quad (15 \text{ bis})$$

condition qui a été obtenue pour les fonctions d'Hermite dans [27] et pour les fonctions trigonométriques dans [28] .

2.- Sous les mêmes hypothèses, et d'une manière tout à fait analogue, on peut montrer que (15) entraîne

$$\lim_{n \rightarrow \infty} E \left[\left(\sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| \right)^2 \right] = 0 \quad (17)$$

et l'on obtient une convergence plus forte que (16). Si f est bornée, (17) peut s'écrire

$$\lim_{n \rightarrow \infty} R(\hat{f}_n, f) = 0$$

R étant la fonction de risque définie au chapitre I quand Θ est l'espace des fonctions continues et bornées sur Ω , muni de la norme de la convergence uniforme.

Vitesse de convergence :

La vitesse de convergence de Q_n vers 0 dépend de celle de la famille sommable $\sum_{i \in I} a_i e_i$ vers f .

Pour mettre en évidence cette dépendance, on peut donner, par exemple, le résultat suivant ⁽¹⁾ :

COROLLAIRE 7 :

On suppose que la famille (e_i) est dénombrable, que les e_i sont bornés uniformément pour $i \in \mathbb{N}$, et que $e_{i_k} = e_k$, $\forall k \in \mathbb{N}$, $\forall n \in \mathbb{N}$.

Alors, si $|a_k| \leq \frac{\alpha}{k^r}$, avec $r > 1$ et $\alpha > 0$, $\forall k \in \mathbb{N}$

et si $q(n) = O(n^{1/2r})$, on a $Q_n = O\left(\frac{1}{\frac{r-1}{r}}\right)$

DÉMONSTRATION.

C'est une conséquence immédiate de l'inégalité

$$Q_n \leq M^4 \left[\frac{q(n)^2}{n} + \left(\sum_{k > q(n)} |a_k| \right)^2 \right]$$

où $M = \max \left(\sup_{\omega \in \Omega, i \in \mathbb{N}} |e_i(\omega)|, 1 \right)$.

3. Convergence uniforme presque sûre.

Nous allons maintenant donner des conditions pour que

$\hat{d}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)|$ tende presque sûrement vers 0 quand n tend vers l'infini.

Bien que la définition 5 du chapitre I ne fasse pas intervenir la mesurabilité de l'estimateur, il semble intéressant d'indiquer une condition suffisante pour que $\hat{d}_n$ soit une variable aléatoire.

(1) Extension de la deuxième partie du théorème 2 de S.C. SCHWARTZ [27].

LEMME 10 :

Soit Ω un espace mesurable, soit X un espace topologique localement compact à base dénombrable, et soit $T(x, \omega)$ une application de $X \times \Omega$ dans $\mathbb{R}^+$. Alors, si $\forall \omega_0 \in \Omega$, $T(x, \omega_0)$ est continue et $\forall x_0 \in X$, $T(x_0, \omega)$ est mesurable, l'application $h(\omega) = \sup_{x \in X} T(x, \omega)$ est mesurable.

DÉMONSTRATION.

Soit K un compact de X et soit d une distance sur X compatible avec sa topologie.

$\forall \omega \in \Omega$, $T(x, \omega)$ est uniformément continue sur K , il existe donc $\alpha_n(\omega) > 0$ tel que

$$d(x, x') < \alpha_n(\omega) \implies |T(x, \omega) - T(x', \omega)| < \frac{1}{n} \quad \forall x, x' \in K; \quad \forall \omega \in \Omega.$$

Soit $\{B(x_i, \alpha_n(\omega))\}$, $i = 1, 2, \dots, \lambda_n$ un recouvrement de K par un nombre fini de Boules ouvertes de rayon $\alpha_n(\omega)$.

L'application $h_{\lambda_n}(\omega) = \sup_{1 \leq i \leq \lambda_n} T(x_i, \omega)$ est mesurable.

Posons $h_K(\omega) = \sup_{x \in K} T(x, \omega)$; on aura

$$h_{\lambda_n}(\omega) \leq h_K(\omega) \leq h_{\lambda_n}(\omega) + \frac{1}{n} \quad \forall \omega \in \Omega$$

donc $\lim_{n \rightarrow \infty} h_{\lambda_n}(\omega) = h_K(\omega)$, $\forall \omega \in \Omega$ ce qui entraîne que h_K est mesurable.

Enfin, soit $(K_p)_{p \in \mathbb{N}}$ une suite croissante de compacts qui recouvre X ; on peut écrire :

$$h(\omega) = \sup_{p \in \mathbb{N}} h_{K_p}(\omega), \quad \forall \omega \in \Omega$$

donc $h(\omega)$ est mesurable.

CONSÉQUENCE :

Si Ω est un espace topologique localement compact à base dénombrable, si f est continue et si e_i est continue, $\forall i \in I$, d_n est mesurable.

Passons maintenant aux théorèmes de convergence. Nous supposons que la famille e_i est dénombrable, que

$$e_{i_k, n} = e_k, \quad \forall k \in \mathbb{N}, \quad \forall n \in \mathbb{N}, \quad \text{que } e_i \text{ est continue,}$$

$\forall i \in \mathbb{N}$, et que f est continue :

THÉORÈME 12 :

Sous les conditions suivantes :

1°) e_i est bornée uniformément pour $i \in \mathbb{N}$.

2°) La série $\sum_{i \in \mathbb{N}} a_i e_i$ converge uniformément vers f .

3°) Il existe une constante positive m telle que

$$\int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \geq m, \quad \forall i \in \mathbb{N}$$

4°) $\lim_{n \rightarrow \infty} q(n) = +\infty$

5°) $\sum_{n=1}^{+\infty} q(n) \exp \left[-\lambda \frac{n}{q(n)^2} \right] < +\infty, \quad \forall \lambda > 0$

$$\text{On a } \lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f(\omega) | = 0$$

DÉMONSTRATION.

Comme $\lim_{n \rightarrow \infty} q(n) = +\infty$, la condition S' est vérifiée, or

$$\sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f(\omega) | \leq \sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f_n(\omega) | + \sup_{\omega \in \Omega} | f_n(\omega) - f(\omega) | \quad (18)$$

il suffit de montrer que

$$\hat{\delta}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f_n(\omega)| \rightarrow 0 \text{ quand } n \rightarrow \infty. \\ \text{p.s.}$$

Soit X une variable aléatoire à valeurs dans Ω et dont la loi a pour densité f ; on a la majoration :

$$\sigma_1^2 = \sigma^2 [e_1(X)] \leq \int_{\Omega} e_1^2(\omega) f(\omega) d\mu(\omega) \leq M^2 \cdot \int_{\Omega} f(\omega) d\mu(\omega) = M^2 \quad (19)$$

$$\forall i \in \mathbf{N}, \text{ où l'on a posé } M = \sup_{i \in \mathbf{N}, \omega \in \Omega} |e_i(\omega)|.$$

D'autre part, il existe $i_0 \in \mathbf{N}$ tel que $i > i_0$ entraîne

$$|a_i|^2 \leq \frac{m}{2}. \text{ D'où compte tenu de } 3^\circ):$$

$$\sigma^2 [e_i(X)] = \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) - a_i^2 \geq \frac{m}{2}, \quad \forall i > i_0 \quad (20)$$

Pour n assez grand, on aura $q(n) > i_0$, donc

$$\hat{\delta}_n = \sup_{\omega \in \Omega} \left| \sum_{i=1}^{q(n)} (\hat{a}_{i_n} - a_i) e_i(\omega) \right| \leq M \sum_{i=1}^{i_0} |\hat{a}_{i_n} - a_i| \\ + M \sum_{i=i_0+1}^{q(n)} |\hat{a}_{i_n} - a_i| \quad (21)$$

$$\text{Comme } \hat{a}_{i_n} = \frac{e_i(\omega_1) + \dots + e_i(\omega_n)}{n} \text{ et } a_i = E(\hat{a}_{i_n}) \text{ la loi}$$

forte des grands nombres montre que $\hat{a}_{i_n} \xrightarrow{\text{p.s.}} a_i^{(1)}$ quand $n \rightarrow \infty$,

$\forall i \in \mathbf{N}$, donc :

$$\lim_{n \rightarrow \infty} \text{p.s.} \left[M \sum_{i=1}^{i_0} |\hat{a}_{i_n} - a_i| \right] = 0$$

(1) Et même presque complètement sûrement.

D'après (21), il suffit de montrer que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| = 0 \quad (22)$$

Pour cela, nous allons appliquer une inégalité de Tchebychev renforcée (1). Il vient :

$$\begin{aligned} \text{Prob} [|e_i(\omega_1) + \dots + e_i(\omega_n) - n a_i| \geq \mu \sigma_i \sqrt{n}] \\ \leq 2 \exp \left[- \frac{\mu^2}{2 \left(1 + \frac{\mu}{2 \sigma_i \sqrt{n}}\right)^2} \right] \end{aligned}$$

$$\text{quel que soit } \mu \text{ tel que } 0 < \mu \leq \frac{\sigma_i \sqrt{n}}{M}.$$

$$\text{Donnons nous } \varepsilon \text{ tel que } 0 < \varepsilon \leq \frac{\sqrt{m}}{\sqrt{2}} \text{ et choisissons } \mu = \frac{\varepsilon \sqrt{n}}{M};$$

$$\text{on a bien } 0 < \mu \leq \frac{\sigma_i \sqrt{n} \sqrt{2}}{M} \text{ puisque}$$

$$0 < \frac{\varepsilon \sqrt{n}}{M} \leq \frac{\sqrt{m}}{\sqrt{2}} \cdot \frac{\sqrt{n}}{M} \leq \sigma_i \frac{\sqrt{n}}{M}$$

Alors, compte tenu de (19) et (20), on peut écrire, $\forall i > i_0$:

$$\text{Prob} [|\hat{a}_{in} - a_i| \geq \varepsilon] \leq \text{Prob} [|\hat{a}_{in} - a_i| \geq \varepsilon \cdot \frac{\sigma_i}{M}]$$

$$\leq 2 \exp \left[- \frac{n \varepsilon^2}{2 M^2 \left(1 + \frac{\varepsilon}{2 \sigma_i}\right)^2} \right]$$

$$\leq 2 \exp \left[- \frac{2 n \varepsilon^2}{9 M^2} \right]$$

$$\text{inégalité valable pour tout } \varepsilon \leq \frac{\sqrt{m}}{\sqrt{2}}.$$

(1) cf. RENYI [30] théorème 1 p. 365, cette inégalité est due à Bernstein.

On a donc

$$\begin{aligned} \text{Prob} \left[\sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \varepsilon \right] &\leq \sum_{i=i_0+1}^{q(n)} \text{Prob} \left[|\hat{a}_{in} - a_i| > \frac{\varepsilon}{q(n) - i_0} \right] \\ &\leq (q(n) - i_0) \cdot 2 \exp \left[-\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{[q(n) - i_0]^2} \right] \\ &\leq 2q(n) \exp \left[-\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{q(n)^2} \right] \end{aligned}$$

d'où (22), d'après la condition 5°), ce qui achève la démonstration.

COROLLAIRE 8 :

Si les conditions 1°), 2°), 3°) et 4°) du théorème 12 sont vérifiées et si $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0$ on a encore :

$$\lim_{n \rightarrow \infty} \text{p.s.} \quad \hat{d}_n = 0$$

DÉMONSTRATION.

Il suffit de montrer que

$$\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0 \implies \sum_{n=1}^{+\infty} q(n) \exp \left[-\lambda \frac{n}{q(n)^2} \right] < +\infty, \quad \forall \lambda > 0.$$

Pour cela posons $\alpha(n) = \frac{q(n)^2}{n} \text{Log } n$, il vient

$$\begin{aligned} q(n) \exp \left[-\lambda \frac{n}{q(n)^2} \right] &= \frac{\alpha(n)^{1/2} n^{1/2}}{(\text{Log } n)^{1/2}} \exp \left[-\lambda \frac{\text{Log } n}{\alpha(n)} \right] \\ &= \left(\frac{\alpha(n)}{\text{Log } n} \right)^{1/2} \cdot \frac{1}{\frac{\lambda}{\alpha(n)} - 1/2} \end{aligned}$$

Donc, pour n assez grand, $q(n) \exp \left[-\lambda \frac{n}{q(n)^2} \right] < \frac{1}{n^2}$ C.Q.F.D.

REMARQUES :

1. On peut comparer la condition $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0$ à la condition

15 bis. Si $q(n) = [An^\alpha]$, $A > 0$, $\alpha > 0$, les deux conditions sont équivalentes.

2. Si Ω est compact et si f est continue et positive sur Ω , la condition 3°) du théorème 12 est vérifiée. En effet, on a alors :

$$\int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \geq \inf_{\omega \in \Omega} f(\omega) \cdot \int_{\Omega} e_i^2(\omega) d\mu(\omega) > 0$$

Si les e_i sont les fonctions trigonométriques, la condition est vérifiée $\forall f \in L^2(\Omega)$. En effet, si i est pair ($i = 2k$), on a, en prenant par exemple $\Omega = [0, 2\pi]$:

$$\begin{aligned} \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) &= \frac{1}{\pi} \int_0^{2\pi} \cos^2 k\omega f(\omega) d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega) d\omega + \frac{1}{2\pi} \int_0^{2\pi} \cos 2k\omega f(\omega) d\omega \\ &\geq \frac{1}{4\pi} \text{ pour } k \text{ assez grand.} \end{aligned}$$

Le calcul est analogue pour i impair.

Nous ne savons pas si la condition 3°) est vérifiée pour n'importe quel système orthonormal, $\forall \Omega$, $\forall \mu$, $\forall f \in L^2(\mu)$.

Nous allons maintenant donner un théorème de convergence uniforme presque sûre, valable pour $\Omega = \mathbf{R}$, et qui ne fait pas intervenir la condition 3°). La méthode de démonstration est analogue à celle de NADARAYA [18].

THÉORÈME 13 :

On suppose que :

1°) Ω est un intervalle de $\mathbf{R}$, fini ou infini et μ est la mesure de Lebesgue sur $\mathbf{R}$.

2°) e_i est la variation bornée, $\forall i \in \mathbf{N}$.

3°) La série $\sum_{i \in \mathbf{N}} a_i e_i$ converge uniformément vers f .

4°) $\lim_{n \rightarrow \infty} q(n) = +\infty$.

Alors, pour que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| = 0$$

il suffit que

$$\sum_{n=1}^{+\infty} \exp \left[-\alpha \frac{n}{M_{q(n)}^2 \cdot V_{q(n)}^2} \right] < +\infty, \quad \forall \alpha > 0 \quad (23)$$

$$\text{avec } M_{q(n)} = \sup_{1 \leq i \leq q(n)} \sup_{\omega \in \Omega} |e_i(\omega)|^{(1)}$$

$$\text{et } V_{q(n)} = \sum_{i=1}^{q(n)} \int |de_i|.$$

DÉMONSTRATION.

Les conditions 3°) et 4°) entraînent la condition S' donc, d'après (18), il suffit de montrer que

$$\hat{\delta}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f_n(\omega)| \rightarrow 0 \text{ p.s. quand } n \rightarrow \infty.$$

Soit F la fonction de répartition associée à f et soit F_n la fonction de répartition empirique. On peut écrire :

$$a_i = \int_{\Omega} e_i(\omega) dF(\omega) \quad \text{et} \quad \hat{a}_{in} = \int_{\Omega} e_i(\omega) dF_n(\omega), \quad \forall i \in \mathbf{N}$$

Donc

$$\begin{aligned} \hat{f}_n(\omega) - f_n(\omega) &= \sum_{i=1}^{q(n)} (\hat{a}_{in} - a_i) e_i(\omega) \\ &= \int_{\Omega} \left[\sum_{i=1}^{q(n)} e_i(t) e_i(\omega) \right] \cdot [dF_n(t) - dF(t)], \quad \forall \omega \in \Omega \end{aligned}$$

(1) e_i étant à variation bornée, $\forall i \in \mathbf{N}$, on a $M_{q(n)} < +\infty$.

comme les e_i sont à variation bornée, on peut intégrer par parties et il vient :

$$\hat{f}_n(\omega) - f_n(\omega) = - \int_{\Omega} [F_n(t) - F(t)] \cdot \left[\sum_{i=1}^{q(n)} e_i(\omega) de_i(t) \right], \quad \forall \omega \in \Omega$$

d'où

$$\hat{\delta}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f_n(\omega)| \leq D_n \cdot M_{q(n)} \cdot V_{q(n)}$$

$$\text{avec } D_n = \sup_{t \in \Omega} |F_n(t) - F(t)|$$

alors, d'après le lemme 6 (Chapitre I) :

$$\text{Prob} [|\hat{\delta}_n| > \varepsilon] \leq \text{Prob} [D_n \cdot M_{q(n)} \cdot V_{q(n)} > \varepsilon]$$

$$\leq C \exp \left[-2 \frac{n \varepsilon^2}{M_{q(n)}^2 \cdot V_{q(n)}^2} \right], \quad \forall \varepsilon > 0 \quad (23 \text{ bis})$$

et (23) permet de conclure.

EXEMPLES :

1. - Si $\Omega = [0, 2\pi]$ et si les e_i sont les fonctions trigonométriques, le théorème 12 s'applique pour f continue à variation bornée et telle que $f(0) = f(2\pi)$ et $q(n) = O(n^r)$, $r < \frac{1}{2}$.

2. - Nous verrons au Chapitre III l'application du théorème 13 dans le cas où les e_i sont les fonctions d'Hermite.

Donnons enfin, sous la forme d'une condition nécessaire et suffisante, un énoncé qui contient une réciproque du théorème 13 :

On suppose que la condition (23) et les hypothèses 1°, 2° et 4° du théorème 13 sont vérifiées et on désigne par $\mathcal{C}$ l'ensemble des densités définies sur (Ω, μ) et admettant, par rapport aux e_i , une série de Fourier uniformément convergente. On a alors le résultat suivant :

THÉORÈME 13 BIS :

Pour que :

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| = 0$$

il faut et il suffit que f appartienne à $\mathcal{C}$.

DÉMONSTRATION.

La condition suffisante n'est autre que le théorème 13.

La condition nécessaire est une conséquence immédiate de l'inégalité :

$$\sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| \leq \hat{\delta}_n + \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)|$$

et du fait que, d'après la démonstration du théorème 13,

$$\lim_{n \rightarrow \infty} \text{p.s.} \hat{\delta}_n = 0 \text{ quelle que soit } f.$$

IV - LOI LIMITE DE L'ESTIMATEUR

Nous supposons, comme au paragraphe précédent, que l'on estime une version de la densité et que $\hat{f}_n$ est défini à l'aide d'une version des e_i . Nous cherchons la loi limite de $\hat{f}_n(\omega_0)$, ω_0 étant un point donné de Ω .

THÉORÈME 14 :

Si

$$1^\circ) M = \sup_{i \in I} \sup_{\omega \in \Omega} |e_i(\omega)| < +\infty$$

2°) il existe $A > 0$ tel que, pour n assez grand :

$$q(n)^{-1} \cdot \int_{\Omega} \left[\sum_{k=1}^{q(n)} e_{i_k, n}(\omega) e_{i_k, n}(\omega_0) \right]^2 f(\omega) d\mu(\omega) \geq A$$

3°) $f_n(\omega_0) \rightarrow f(\omega_0)$ quand $n \rightarrow +\infty$

4°) Il existe $\alpha > 0$ tel que $\lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0$

5°) $\lim_{n \rightarrow \infty} q(n) = +\infty$

Alors $\frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]}$ converge en loi vers une variable aléatoire

dont la loi est normale de moyenne nulle et d'écart type 1.

DÉMONSTRATION.

On peut écrire $\hat{f}_n(\omega_0)$ sous la forme d'une somme de variables aléatoires indépendantes et équidistribuées, en effet :

$$\begin{aligned}\hat{f}_n(\omega_0) &= \sum_{k=1}^{q(n)} \frac{1}{n} [e_{i_k, n}(\omega_1) + \dots + e_{i_k, n}(\omega_n)] e_{i_k, n}(\omega_0) \\ &= \sum_{k=1}^n y_{kn}\end{aligned}$$

$$\text{avec } y_{kn} = \frac{1}{n} \sum_{j=1}^{q(n)} e_{i_j, n}(\omega_k) e_{i_j, n}(\omega_0) \quad (24)$$

$$\text{Donc } \frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]} = \sum_{k=1}^n z_{kn}$$

$$\text{avec } z_{kn} = \frac{y_{kn} - E(y_{kn})}{\sqrt{n} \sigma[y_{kn}]} \quad k = 1, \dots, n.$$

Nous allons montrer que l'on peut appliquer le théorème central limite (cf LOEVE [31] p. 316) aux variables z_{kn} . Pour cela, il suffit de montrer que :

$$\text{a) } P_n(\varepsilon) = \sum_{k=1}^n \text{Prob}[|z_{kn}| > \varepsilon] \rightarrow 0 \text{ quand } n \rightarrow +\infty, \forall \varepsilon > 0$$

b) Il existe $\tau > 0$ tel que, quand $n \rightarrow +\infty$

$$\sum_{k=1}^n \sigma^2_{kn}(z_{kn}(\tau)) \rightarrow 1 \text{ et } \sum_{k=1}^n E[z_{kn}(\tau)] \rightarrow 0$$

$z_{kn}(\tau)$ désignant la variable aléatoire z_{kn} tronquée à τ .

DÉMONSTRATION DE a).

On peut écrire :

$$P_n(\varepsilon) = n \text{ Prob}(|z_{1n}| > \varepsilon) \leq \frac{n}{\varepsilon^{2+2\alpha}} \cdot \frac{E[|y_{\ell n} - E(y_{\ell n})|^{2+2\alpha}]}{n^{1+\alpha} (\sigma[y_{\ell n}])^{2+2\alpha}} \quad (25)$$

Or, d'après (24) et la condition 1°)

$$|y_{\ell n}| \leq M^2 \frac{q(n)}{n} \quad (26)$$

et d'après la condition 3°), il existe $\delta > 0$ tel que, pour n assez grand :

$$|E(y_{\ell n})| = \frac{f_n(\omega_0)}{n} \leq \frac{f(\omega_0) + \delta}{n} \quad (27)$$

Donc, d'après 5°), on aura, pour n assez grand :

$$|y_{\ell n} - E(y_{\ell n})| \leq 2M^2 \frac{q(n)}{n} \quad (28)$$

D'autre part, en tenant compte de 2°), on obtient

$$\begin{aligned} E(y_{\ell n}^2) &= \frac{1}{n^2} \int_{\Omega} \left[\sum_{k=1}^{q(n)} e_{i_{k,n}}(\omega) e_{i_{k,n}}(\omega_0) \right]^2 f(\omega) d\mu(\omega) \\ &\geq A \frac{q(n)}{n^2} \text{ pour } n \text{ assez grand} \end{aligned} \quad (29)$$

(27) et (28) entraînent, pour n assez grand :

$$\sigma^2(y_{\ell n}) = E(y_{\ell n}^2) - [E(y_{\ell n})]^2 \geq \frac{A}{2} \cdot \frac{q(n)}{n^2} \quad (30)$$

En reportant dans (25), on trouve :

$$\begin{aligned} P_n(\varepsilon) &\leq \frac{1}{\varepsilon^{2+2\alpha}} \cdot \frac{1}{n^\alpha} (2M^2)^{2+2\alpha} \cdot \frac{q(n)^{2+2\alpha}}{n^{2+2\alpha}} \cdot \left(\frac{2}{A}\right)^{1+\alpha} \cdot \frac{n^{2+2\alpha}}{q(n)^{1+\alpha}} \\ \text{soit } P_n(\varepsilon) &\leq K(\varepsilon) \frac{q(n)^{1+\alpha}}{n^\alpha} \end{aligned} \quad (31)$$

avec $K(\varepsilon) = \left(\frac{2\sqrt{2} M^2}{\varepsilon \sqrt{A}} \right)^{2+2\alpha}$

d'où a) d'après 4°).

DÉMONSTRATION DE b).

D'après (28) et (30), on peut écrire

$$|z_{\ell n}| \leq \frac{|y_{\ell n} - E(y_{\ell n})|}{\sqrt{n} \sigma(y_{\ell n})} \leq \frac{1}{\sqrt{n}} \cdot 2M^2 \frac{q(n)}{n} \cdot \sqrt{\frac{2}{A}} \frac{n}{\sqrt{q(n)}}$$

soit $|z_{\ell n}| \leq \frac{2\sqrt{2} M^2}{\sqrt{A}} \cdot \sqrt{\frac{q(n)}{n}}$

Or 4°) entraîne $\sqrt{\frac{q(n)}{n}} \rightarrow 0$ puisque

$$\left(\sqrt{\frac{q(n)}{n}} \right)^{2+2\alpha} = \frac{1}{n} \cdot \frac{q(n)^{1+\alpha}}{n^\alpha}$$

donc, $\forall \tau > 0$, $z_{\ell n}(\tau) = z_{\ell n}$ pour n assez grand,

et comme $\sigma(z_{\ell n}) = \frac{1}{\sqrt{n}}$ et $E(z_{\ell n}) = 0$ b) est démontré.

CONCLUSION.

Le théorème limite central s'applique et

$$\frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]} \xrightarrow{\mathcal{L}} \mathcal{N}_{(0,1)} \quad \text{C.Q.F.D.}$$

EXEMPLE :

Si les e_i sont les fonctions trigonométriques sur $[0, 2\pi]$, si f est à variation bornée, telle que $\inf_{0 \leq \omega \leq 2\pi} f(\omega) > 0$ et continue au point ω_0 ,

enfin si $q(n) = \bigcirc (n^\beta)$, $0 < \beta < 1$, le théorème 14 s'applique
 $\forall \omega_0 \in]0, 2\pi[$.

V - APPLICATIONS

1. Estimation du mode .

COROLLAIRE 9 :

Pour $\Omega = \mathbf{R}^p$ et $\mu = \lambda^p$ (mesure de Lebesgue sur $\mathbf{R}^p$), si e_i est continue et telle que $\lim_{\|\omega\|=\infty} e_i(\omega) = 0$, $\forall i \in \mathbf{N}$; si f est uniformément

continue et unimodale et si $\hat{m}_n$ est une solution de l'équation

$$\hat{f}_n(\hat{m}_n) = \sup_{\omega \in \mathbf{R}^p} \hat{f}_n(\omega),$$

alors $\hat{m}_n$ est un estimateur presque sûrement convergent du mode m de f quand les hypothèses du théorème 12 ⁽¹⁾ sont vérifiées.

DÉMONSTRATION.

La démonstration est tout à fait analogue à celle du théorème 3A de Parzen [29] et du théorème 2 de Nadaraya [18]. On a, en effet, les propriétés suivantes :

$$1^\circ) \quad \forall \varepsilon > 0, \quad \alpha > 0 \text{ tel que } \|m - \omega\| \geq \varepsilon \Rightarrow |f(m) - f(\omega)| \geq \alpha \quad (2)$$

$$2^\circ) \quad |f(\hat{m}_n) - f(m)| \leq 2 \sup_{\omega \in \mathbf{R}^p} |\hat{f}_n(\omega) - f(\omega)| \quad (2)$$

donc, $\forall \varepsilon > 0$

$$\text{Prob} [\|\hat{m}_n - m\| \geq \varepsilon] \leq \text{Prob} [|f(\hat{m}_n) - f(m)| \geq \alpha]$$

$$\leq \text{Prob} \left[\sup_{\omega \in \mathbf{R}} |\hat{f}_n(\omega) - f(\omega)| \geq \frac{\alpha}{2} \right]$$

(1) Resp. du théorème 13 si $p = 1$.

(2) cf PARZEN p. 1072.

d'où le résultat puisque

$$\sum_{n=1}^{\infty} \text{Prob} \left[\sup_{\omega \in \mathbf{R}} \left| \hat{f}_n(\omega) - f(\omega) \right| \geq \frac{\alpha}{2} \right] < +\infty.$$

2. Estimation de la fonction de répartition.

Ω étant un intervalle compact de $\mathbf{R}$ muni de la mesure de Lebesgue, on peut définir un estimateur de la fonction de répartition F , analogue à $\hat{f}_n$, en posant :

$$\hat{F}_n(\omega; \omega_1, \dots, \omega_n) = \sum_{j=1}^{q(n)} \left[\int_{\Omega} e_j(t) F_n^*(t; \omega_1, \dots, \omega_n) dt \right] e_j(\omega) \quad (32)$$

$\forall n \in \mathbf{N}$, $\forall (\omega; \omega_1, \dots, \omega_n) \in \Omega^{n+1}$, où F_n^* désigne la fonction de répartition empirique et où les e_j forment une base orthonormale de $\mathcal{L}^2(\Omega)$ ⁽¹⁾.

COROLLAIRE 10 :

1°) Si $\lim_{n \rightarrow \infty} q(n) = +\infty$, $\hat{F}_n$ converge vers F en moyenne quadratique intégrée.

2°) Si les e_j sont bornés uniformément pour $j \in \mathbf{N}$,

si $\sum_{j \in \mathbf{N}} \left[\int_{\Omega} e_j F \right] e_j$ converge uniformément vers F et

si $\sum_{n=1}^{+\infty} \exp \left[-\gamma \frac{q(n)}{n} \right] < +\infty$, $\forall \gamma > 0$,

alors $\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} \left| \hat{F}_n(\omega) - F(\omega) \right| = 0$

DÉMONSTRATION.

1°) Posons $\Omega = [a, b]$ ($a < b$), $\hat{A}_{jn} = \int_a^b e_j F_n^*$ et

(1) Le cas des fonctions trigonométriques a été abordé dans [28].

$$A_j = \int_a^b e_j F. \text{ Alors :}$$

$$J_n^* = E \int_a^b (F_n^* - F)^2 = \sum_{j=1}^{+\infty} \text{Var } \hat{A}_{jn}$$

$$\text{et } \hat{J}_n = E \int_a^b (\hat{F}_n - F)^2 = \sum_{j=1}^{q(n)} \text{Var } \hat{A}_{jn} + \sum_{j > q(n)} A_j^2$$

$$\text{D'où, en remarquant que } E(\hat{A}_{jn}) = A_{jn} :$$

$$\begin{aligned} J_n &= J_n^* + \sum_{j > q(n)} [A_j^2 - \text{Var } \hat{A}_{jn}] \\ &= 2 \sum_{j > q(n)} A_j^2 - \sum_{j > q(n)} E(\hat{A}_{jn}^2) + J_n^* \end{aligned} \quad (33)$$

$$\text{Or, } J_n^* = \frac{1}{n} \int_a^b F(1-F)$$

$$\sum_{j=1}^{\infty} A_j^2 = \int_a^b F^2$$

$$\text{et } \sum_{j=1}^{\infty} E(\hat{A}_{jn}^2) = \int_a^b F_n^{*2} < b-a$$

$$\text{D'où le résultat puisque } q(n) \rightarrow \infty .$$

2°) La démonstration est analogue à celle du théorème 13 :

$$\text{On pose } E(\hat{F}_n) = F_n \text{ et l'on écrit :}$$

$$\begin{aligned} \hat{F}_n(\omega) - F_n(\omega) &= \sum_{j=1}^{q(n)} (\hat{A}_{jn} - A_{jn}) e_j(\omega) \\ &= \int \left[\sum_{j=1}^{q(n)} e_j(t) e_j(\omega) \right] \cdot [F_n^*(t) - F(t)] dt \end{aligned}$$

$$\forall \omega \in [a, b]$$

d'où, d'après l'inégalité de Schwarz :

$$\begin{aligned} \sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F_n(\omega)| &\leq \left[\int_a^b \left(\sum_{j=1}^{q(n)} e_j(t) e_j(\omega) \right)^2 dt \right]^{1/2} \\ &= \left[\int_a^b [F_n^*(t) - F(t)]^2 dt \right]^{1/2} \\ &\leq M q(n)^{1/2} (b-a)^{1/2} \sup_{\omega \in \Omega} |F_n^*(\omega) - F(\omega)| \end{aligned}$$

$$\text{avec } M = \sup_{\omega \in [a,b]} \sup_{j \in \mathbf{N}} |e_j(\omega)|.$$

et l'on conclut en utilisant le lemme 6 (Chapitre I) et l'inégalité

$$\begin{aligned} \sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F(\omega)| &\leq \sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F_n(\omega)| \\ &\quad + \sup_{\omega \in [a,b]} |F_n(\omega) - F(\omega)| \end{aligned}$$

REMARQUE :

Si les e_j sont les fonctions trigonométriques sur $[a,b]$, $\hat{F}_n$ est meilleur (au sens du MISE) que F_n^* pour un choix convenable de $q(n)$ (cf [28] théorème 5, p.950). Ceci n'est pas en contradiction avec le théorème 9 (Chapitre I) car $\hat{F}_n$ n'est pas, en général, un estimateur sans biais de F .

III

ESTIMATION DES DÉRIVÉES DE LA DENSITÉ

I - GÉNÉRALITÉS

1. Introduction.

Le premier article sur l'estimation des dérivées de la densité semble être celui de P.K. BHATTACHARYA [33] (1967), ses résultats ont été complétés par SCHUSTER [14] (1969). Ces deux auteurs estiment la $k^{\text{ème}}$ dérivée de la densité à l'aide de la $k^{\text{ème}}$ dérivée de l'estimateur de la densité défini par la méthode du noyau. SCHUSTER a obtenu des conditions de convergence uniforme presque sûre.

Dans ce chapitre, nous étudierons des méthodes liées à celle des fonctions orthogonales. Les démonstrations étant analogues à celles du chapitre précédent, nous les exposerons brièvement; nous appliquerons ensuite les résultats obtenus à certains cas particuliers, notamment au cas des fonctions d'Hermite ⁽¹⁾.

2. Méthodes d'estimation.

Dans toute la suite du chapitre, nous supposerons que Ω est un intervalle fini ou infini de $\mathbf{R}$ et que μ est la mesure de Lebesgue sur Ω .

Remarquons que, dans le problème qui nous intéresse ici, on peut envisager deux points de vue :

- 1^{er} point de vue: sachant que f est k fois dérivable, on cherche à estimer directement $f^{(k)}$ ($k^{\text{ème}}$ dérivée de f).

- 2^{ème} point de vue: sachant que f est k fois dérivable, on cherche à obtenir un estimateur de f qui converge suivant une topologie plus fine que celle de la convergence uniforme (en particulier la topologie de la convergence uniforme des dérivées jusqu'à l'ordre k).

Il est clair que l'estimateur choisi pourra dépendre du point de vue adopté et du mode de convergence utilisé. Aussi allons-nous définir deux méthodes pour l'estimation de $f^{(k)}$.

(1) La plupart des résultats qui suivent sont dans notre note aux C.R. [34] du 24 novembre 1969, cf aussi [48].

1ère méthode :

Posons $\Omega = [a, b]$, $a < b$, $a \in \overline{\mathbf{R}}$, $b \in \overline{\mathbf{R}}$ et désignons par $\mathcal{L}^2(\Omega, k)$ l'espace des fonctions réelles définies sur Ω , k fois dérivables et de carré intégrable ainsi que leurs k dérivées.

Soit $(e_i)_{i \in \mathbf{N}}$ une base orthonormale de $\mathcal{L}^2(\Omega)$ telle que $e_i \in \mathcal{L}^2(\Omega, k)$ et telle que

$$\lim_{x \rightarrow a(+)} e_i^{(h)}(x) f^{(k-h)}(x) = \lim_{x \rightarrow b(-)} e_i^{(h)}(x) f^{(k-h)}(x) \quad (1)$$

$$\forall i \in \mathbf{N}, \quad \forall h = 0, 1, \dots, k.$$

On peut alors estimer $f^{(k)}$ par la méthode des fonctions orthogonales, ce qui revient à estimer les coefficients

$$a_{ik} = \int_{\Omega} f^{(k)}(x) e_i(x) dx, \quad i \in \mathbf{N}$$

Or, en intégrant k fois par parties et en tenant compte de (1), il vient

$$a_{ik} = (-1)^k \int_{\Omega} e_i^{(k)}(x) f(x) dx, \quad \forall i \in \mathbf{N}$$

Il semble donc naturel de choisir comme estimateur de a_{ik} :

$$\hat{a}_{ik,n} = (-1)^k \int_{\Omega} e_i^{(k)}(x) dF_n(x), \quad \forall i \in \mathbf{N} \quad (2)$$

où F_n désigne la fonction de répartition empirique associée à un échantillon de taille n .

L'estimateur sera donc de la forme :

$$\hat{f}_{n,k}(x) = \sum_{i=1}^{q(n)} \hat{a}_{ik,n} e_i(x), \quad \forall x \in \Omega, \quad \forall n \in \mathbf{N} \quad (3)$$

avec $\lim_{n \rightarrow \infty} q(n) = +\infty$.

2ème méthode :

Soit $\mathcal{E}_k(\Omega)$ l'espace des fonctions réelles, définies et bornées sur Ω et possédant k dérivées continues et bornées.

Soit $(e_i)_{i \in \mathbb{N}}$ une base orthonormale de $\mathcal{L}^2(\Omega)$ telle que $e_i \in \mathcal{E}_k(\Omega)$, $\forall i \in \mathbb{N}$ et soit f une densité appartenant à $\mathcal{E}_k(\Omega)$.

$\hat{f}_n$ désignant l'estimateur de f étudié au chapitre précédent, sa dérivée $k^{\text{ème}}$, soit $\hat{f}_n^{(k)}$, est un estimateur de $f^{(k)}$. L'estimateur est donc de la forme :

$$\hat{f}_n^{(k)}(x) = \sum_{i=1}^{q(n)} \hat{a}_{in} \cdot e_i^{(k)}(x), \quad \forall x \in \Omega, \quad \forall n \in \mathbb{N} \quad (4)$$

$$\text{avec } \hat{a}_{in} = \int_{\Omega} e_i(x) dF_n(x), \quad \forall i \in \mathbb{N}, \quad \forall n \in \mathbb{N} \quad (5)$$

et $\lim_{n \rightarrow \infty} q(n) = +\infty$.

II - THÉORÈMES ASYMPTOTIQUES

1. Convergence de $\hat{f}_{n,k}$.

THÉORÈME 15 :

1°) Pour que $\hat{f}_{n,k}$ converge vers $f^{(k)}$ en moyenne quadratique intégrée; il faut et il suffit que

$$\lim_{n \rightarrow \infty} \int_{\Omega} \left[\frac{1}{n} \sum_{i=1}^{q(n)} e_i^{(k)2}(x) \right] f(x) dx = 0 \quad (6)$$

2°) Si e_i est bornée et si $e_i^{(k)}$ est à variation bornée, $\forall i \in \mathbb{N}$; si, de plus la série $\sum_{i \in \mathbb{N}} a_{ik} e_i$ converge uniformément vers $f^{(k)}$, alors, pour que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in \Omega} |\hat{f}_{n,k}(x) - f^{(k)}(x)| = 0$$

il suffit que

$$\sum_{n=1}^{+\infty} \exp \left[-\alpha \frac{n}{M_{q(n)}^2 \cdot V_{q(n),k}^2} \right] < +\infty, \quad \forall \alpha > 0 \quad (7)$$

$$\text{avec } M_{q(n)} = \sup_{1 \leq i \leq q(n)} \sup_{x \in \Omega} |e_i(x)|$$

$$\text{et } V_{q(n),k} = \sum_{i=1}^{q(n)} \int_{\Omega} |d e_i^{(k)}|$$

DÉMONSTRATION.

La démonstration de la 1ère partie est tout à fait analogue à celle du théorème 10 ; en effet, le MISE s'écrit

$$J_n = \int_{\Omega} \frac{1}{n} \left[\sum_{i=1}^{q(n)} e_i^{(k)^2}(x) \right] \cdot f(x) dx - \frac{n+1}{n} \sum_{i=1}^{q(n)} a_{ik}^2 + \sum_{i \in N} a_{ik}^2$$

ce qui permet de conclure facilement.

De même, la démonstration de la 2ème parties s'effectue comme celle du théorème 13 : une intégration par parties permet d'écrire l'inégalité :

$$\sup_{x \in \Omega} |\hat{f}_{n,k}(x) - E \hat{f}_{n,k}(x)| \leq D_n \cdot M_{q(n)} \cdot V_{q(n),k}$$

et l'on conclut à l'aide du lemme 6 .

2. Convergence de $\hat{f}_n^{(k)}$.

Munissons $\mathcal{E}_k(\Omega)$ de la norme suivante :

$$\|\varphi\|_k = \sup_{0 \leq \ell \leq k} \sup_{x \in \Omega} |\varphi^{(\ell)}(x)|, \quad \forall \varphi \in \mathcal{E}_k(\Omega)$$

et désignons par C_k l'ensemble des fonctions $\varphi \in \mathcal{E}_k(\Omega)$ telles que

$$\lim_{n \rightarrow \infty} \left\| \sum_{i=1}^n \left[\int_{\Omega} e_i \varphi \right] e_i - \varphi \right\|_k = 0.$$

On a alors les résultats suivants :

THÉORÈME 16 :

On suppose qu'il existe une constante positive m telle que

$$\int_{\Omega} e_i^2(x) f(x) dx \geq m, \quad \forall i \in \mathbb{N}$$

$$\text{que } M = \sup_{i \in \mathbb{N}} \cdot \sup_{x \in \Omega} |e_i(x)| < +\infty$$

et que

$$\sum_{n=1}^{+\infty} q(n) \exp \left[-\lambda \frac{n}{M_{q(n),k}^2 \cdot q(n)^2} \right] < +\infty, \quad \forall \lambda > 0 \quad (8)$$

$$\text{où } M_{q(n),k} = \sup_{1 \leq i \leq q(n)} \|e_i\|_k.$$

Alors pour que

$$\lim_{n \rightarrow \infty} \text{p.s. } \|\hat{f}_n - f\|_k = 0$$

il faut et il suffit que f appartienne à C_k .

DÉMONSTRATION.

Nous allons procéder de la même façon que pour le théorème 12.

Montrons d'abord que :

$$\|\hat{f}_n - E \hat{f}_n\|_k \xrightarrow{\text{p.s.}} 0, \quad \forall f \in \mathcal{E}_k(\Omega)$$

Pour cela, il suffit de remarquer que les relations obtenues dans la démonstration du théorème 12 restent valables, c'est-à-dire qu'il existe un entier i_0 tel que :

$$\text{Prob} \left[\sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \varepsilon \right] \leq 2q(n) \exp \left[-\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{q(n)^2} \right] \quad (9)$$

quel que soit $\varepsilon \in]0, \frac{\sqrt{m}}{\sqrt{2}}[$.

Or,

$$\begin{aligned} \|\hat{f}_n - E\hat{f}_n\|_k &\leq \left[\sup_{1 \leq i \leq i_0} \|e_i\|_k \right] \cdot \sum_{i=1}^{i_0} |\hat{a}_{in} - a_i| \\ &\quad + M_{q(n),k} \cdot \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| \end{aligned}$$

et comme $\sum_{i=1}^{i_0} |\hat{a}_{in} - a_i| \xrightarrow[p.s.]{} 0$, il suffit de montrer que

$$M_{q(n),k} \cdot \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| \xrightarrow[p.s.]{} 0. \text{ Mais, compte tenu de (9), on peut}$$

écrire :

$$\text{Prob} \left[\sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \frac{\varepsilon}{M_{q(n),k}} \right] \leq 2q(n) \exp \left[-\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{M_{q(n),k}^2 \cdot q(n)^2} \right]$$

$$\forall \varepsilon \in]0, \frac{M_{q(n),k} \cdot \sqrt{m}}{2} [\text{ donc, a fortiori,}$$

$$\forall \varepsilon \in]0, \frac{\|e_1\|_k \cdot \sqrt{m}}{\sqrt{2}} [.$$

La relation (8) permet alors de conclure.

La démonstration du théorème s'achève facilement en utilisant les inégalités :

$$\|\hat{f}_n - f\|_k \leq \|\hat{f}_n - E\hat{f}_n\|_k + \|E\hat{f}_n - f\|_k$$

$$\text{et} \quad \|E\hat{f}_n - f\|_k \leq \|E\hat{f}_n - \hat{f}_n\|_k + \|\hat{f}_n - f\|_k$$

THÉORÈME 17 :

Si e_i est à variation bornée, $\forall i \in \mathbb{N}$; si

$$\mu_{q(n),1} = \sup_{1 \leq i \leq q(n)} \sup_{x \in \Omega} |e_i^{(\ell)}(x)| < +\infty, \quad \forall n \in \mathbb{N}$$

$\forall \ell = 0, 1, \dots, k$

et si

$$\sum_{n=1}^{+\infty} \exp \left[-\alpha \frac{n}{\mu_{q(n), \ell}^2 \cdot V_{q(n)}^2} \right] < +\infty, \quad \forall \alpha > 0$$

$$\forall \ell = 0, 1, \dots, k$$

$$où \quad V_{q(n)} = \sum_{i=1}^{q(n)} \int_{\Omega} |d e_i(x)|$$

$$\text{Alors, pour que } \lim_{n \rightarrow \infty} \text{p.s. } \|\hat{f}_n - f\|_k = 0$$

Il faut et il suffit que f appartienne à C_k .

DÉMONSTRATION.

Ce théorème est une généralisation du théorème 13 bis et la méthode de démonstration est analogue :

Il est clair que tout revient à montrer que

$$\|\hat{f}_n - E \hat{f}_n\|_k \xrightarrow{\text{p.s.}} 0, \quad \text{donc que } \sup_{x \in \Omega} |\hat{f}_n^{(\ell)}(x) - E \hat{f}_n^{(\ell)}(x)| \xrightarrow{\text{p.s.}} 0$$

pour $\ell = 0, 1, \dots, k$.

Or, on a :

$$\sup_{x \in \Omega} |\hat{f}_n^{(\ell)}(x) - E \hat{f}_n^{(\ell)}(x)| < D_n \cdot \mu_{q(n), \ell} \cdot V_{q(n)}$$

$$\forall \ell = 0, 1, \dots, k$$

ce qui permet d'achever facilement en utilisant de nouveau le lemme 6.

3. Loi limite de $\hat{f}_n^{(k)}$.

THÉORÈME 18 :

Si

$$1^o) \mu_{q(n), k} = O(q(n)^k)$$

2°) Il existe $A > 0$ tel que, pour n assez grand :

$$q(n)^{-2k-1} \int_{\Omega} \left[\sum_{i=1}^{q(n)} e_i(t) e_i^{(k)}(x) \right]^2 \cdot f(t) dt \geq A$$

$$3^\circ) E \hat{f}_n^{(k)}(x) \rightarrow f^{(k)}(x) \quad \text{quand } n \rightarrow +\infty$$

$$4^\circ) \text{ Il existe } \alpha > 0 \text{ tel que } \lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0$$

Alors

$$\frac{\hat{f}_n^{(k)}(x) - E \hat{f}_n^{(k)}(x)}{\sigma[\hat{f}_n^{(k)}(x)]} \underset{\mathcal{L}}{\mathcal{N}}(0,1)$$

DÉMONSTRATION.

Ce théorème, qui est une généralisation du théorème 14, se démontre de la même façon que ce dernier :

On écrit $\hat{f}_n^{(k)}(x)$ sous la forme

$$\hat{f}_n^{(k)}(x) = \sum_{j=1}^n y_{jn}^{(k)} \quad \text{avec} \quad y_{jn}^{(k)} = \frac{1}{n} \sum_{i=1}^{q(n)} e_i^{(k)}(x) \cdot e_i(x_j),$$

$j = 1, \dots, n$.

Les hypothèses faites entraînent :

$$|y_{1n}^{(k)} - E(y_{1n}^{(k)})| = O\left(\frac{q(n)^{k+1}}{n}\right)$$

et

$$\frac{1}{\sigma^2(y_{1n}^{(k)})} = O\left(\frac{n^2}{q(n)^{2k+1}}\right)$$

l'inégalité

$$\text{Prob} \left[\frac{|y_{1n} - E(y_{1n}^{(k)})|}{\sqrt{n} \sigma(y_{1n}^{(k)})} \geq \varepsilon \right] \leq \frac{E[|y_{1n}^{(k)} - E(y_{1n}^{(k)})|^{2+2\alpha}]}{\varepsilon^{2+2\alpha} n^{1+\alpha} \sigma(y_{1n}^{(k)})^{2+2\alpha}}$$

et le théorème limite central permet alors de conclure.

REMARQUE :

La condition imposée à $q(n)$ est indépendante de $k \in \mathbf{N}$. Pour assurer cette condition, on peut choisir, par exemple, $q(n) = n^\gamma$, $0 < \gamma < 1$; la condition est alors vérifiée pour

$$\alpha > \frac{\gamma}{1 - \gamma}.$$

III - ETUDE DE CAS PARTICULIERS

1. Fonctions trigonométriques.

On suppose que Ω est un intervalle fini $[a, b]$ et que les e_i sont les fonctions trigonométriques sur $[a, b]$.

En appliquant les théorèmes du § II, on obtient les résultats suivants :

COROLLAIRE 11 :

Si les e_i sont les fonctions trigonométriques sur $[a, b]$

si $f \in \mathcal{L}^2([a, b], k)$ et si

$$f^{(\ell)}(a) = f^{(\ell)}(b) \quad \forall \ell = 0, 1, \dots, k \quad (10)$$

Alors,

1°) $\hat{f}_{n,k}$ tend vers $f^{(k)}$ en moyenne quadratique intégrée, si et seulement

$$\text{si} \quad \lim_{n \rightarrow \infty} \frac{q(n)^{2k+1}}{n} = 0.$$

2°) Si $f^{(k)}$ est continue et à variation bornée, et si

$$\sum_{n=1}^{\infty} \exp \left[-\alpha \frac{n}{q(n)^{2k+2}} \right] < +\infty, \quad \forall \alpha > 0$$

$$\text{alors} \quad \sup_{a \leq x \leq b} |\hat{f}_{n,k}(x) - f^{(k)}(x)| \xrightarrow{\text{p.s.}} 0$$

quand $n \rightarrow \infty$.

3°) Si $f \in \mathcal{E}_{k+2}([a, b])$, si $f^{(k+1)}(a) = f^{(k+1)}(b)$;

$i = 1, 2$ si $\inf_{a \leq x \leq b} f(x) > 0$ et s'il existe $\alpha > 0$ tel que

$$\lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0 \quad \text{alors} \quad \frac{\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)}{\sigma[\hat{f}_n^{(k)}(x)]} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

4°) Si $f \in \mathcal{E}_{k+2}([a, b])$, si $f^{(k+1)}(a) = f^{(k+1)}(b)$

$i = 1, 2$ et si

$$\sum_{n=1}^{+\infty} \exp \left[-\beta \frac{n}{q(n)^{2k+2}} \right] < +\infty, \quad \forall \beta > 0 \quad (11)$$

$$\text{alors} \quad \lim_{n \rightarrow \infty} \text{p.s.} \quad \|\hat{f}_n - f\|_k = 0.$$

DÉMONSTRATION.

Des calculs trigonométriques élémentaires montrent que :

a) $\sum_{i=1}^{q(n)} e_i^{(k)}(x)$ est de l'ordre de $q(n)^{2k+1}$, uniformément pour

$$x \in [a, b].$$

b) $V_{q(n), \ell}$ est de l'ordre de $q(n)^{\ell+1}$, $\forall \ell = 0, 1, \dots, k$.

c) $M_{q(n), k}$ est de l'ordre de $q(n)^k$.

d) $\mu_{q(n), \ell}$ est de l'ordre de $q(n)^\ell$, $\forall \ell = 0, 1, \dots, k$.

e) $\int_a^b \left[\sum_{i=1}^{q(n)} e_i(t) e_i^{(k)}(x) \right]^2 f(t) dt$ est au moins de l'ordre de

$$q(n)^{2k+1} \quad \text{si} \quad \inf_{a \leq x \leq b} f(x) > 0.$$

Dans ces conditions :

DÉMONSTRATION DU 1°) :

C'est une conséquence immédiate de a) et de la première partie du théorème 15 .

DÉMONSTRATION DU 2°) :

Comme $f^{(k)}$ est continue à variation bornée, $E \hat{f}_{n,k}(x) \rightarrow f^{(k)}(x)$ uniformément sur $[a,b]$ et l'on peut, en tenant compte de b), appliquer la deuxième partie du théorème 15 .

DÉMONSTRATIONS DU 3°) ET DU 4°) :

Des hypothèses faites sur f , on déduit que les coefficients de Fourier de f sont, au plus, de l'ordre de $\frac{1}{n^{k+2}}$ ce qui implique la convergence uniforme de $E \hat{f}_n^{(\ell)}$ vers $f^{(\ell)}$ pour $\ell = 0, 1, \dots, k$. Alors de d), e) et du théorème 18, on tire 3°). Enfin de b), d) et du théorème 17, on tire 4°).

REMARQUES :

1 - Les conditions (10) présentent des inconvénients pour les applications. On pourrait éviter ces inconvénients en estimant f par son développement en série par rapport à un système orthonormal sur $[a,b]$ tel que $e_i(a) = e_i(b) = 0$, $\forall i \in \mathbb{N}$.

2 - On pourrait, pour obtenir une convergence uniforme presque sûre de $\|\hat{f}_n - f\|_k$, utiliser le théorème 16 ; mais la condition imposée à $q(n)$ serait plus forte que (11). En effet, d'après c), on aurait la condition :

$$\sum_{n=1}^{\infty} q(n) \exp \left[-\beta \frac{n}{q(n)^{2k+2}} \right] < +\infty, \quad \forall \beta > 0.$$

COROLLAIRE 12 :

Si f est indéfiniment dérivable et telle que

$f^{(k)}(a) = f^{(k)}(b)$, $\forall k \in \mathbb{N}$; si les e_i sont les fonctions trigonomé-

triques sur $[a, b]$ et si $q(n) = (\text{Log } n)$, alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{a \leq x \leq b} |\hat{f}_n^{(k)}(x) - f^{(k)}(x)| = 0, \quad \forall k \in \mathbb{N} \quad (12)$$

DÉMONSTRATION.

Il est clair que l'on peut appliquer le 4°) du corollaire 11, $\forall k \in \mathbb{N}$, d'où le résultat.

REMARQUE :

On peut donner le même énoncé pour $\hat{f}_{n,k}$.

2. Fonctions d'Hermite.

Sur $L^2(\mathbb{R})$ le système orthonormal le plus utilisé est construit avec les fonctions d'Hermite h_j ($j = 0, 1, \dots$). Elles sont définies de la façon suivante :

$$h_j(x) = (2^j j! \pi^{1/2})^{-1/2} \cdot H_j(x) e^{-x^2/2} \quad j = 0, 1, \dots; \quad x \in \mathbb{R}$$

$$\text{avec } H_j(x) = e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}) \quad j = 0, 1, \dots; \quad x \in \mathbb{R}$$

Les H_j sont les polynômes d'Hermite.

L'application des résultats du § II nécessite la démonstration de quelques lemmes :

LEMME 11 (S.C. SCHWARTZ)

Si $e^{-x^2} \frac{d^k}{dx^k} [e^{-x^2/2} f(x)]$ est de carré intégrable

$$\text{alors } |a_j| = \left| \int_{\mathbb{R}} f(x) h_j(x) dx \right| = O(j^{-k/2})$$

DÉMONSTRATION.

Voir S.C. SCHWARTZ [27] p. 1262.

LEMME 12 :

Pour $j \geq 1$ et $k \geq 0$, on a l'inégalité

$$\sup_{x \in \mathbf{R}} |h_j^{(k)}(x)| \leq c [2(j+k)]^{k/2} \quad (13)$$

où c est une constante indépendante de j et k .

DÉMONSTRATION.

Nous allons procéder par récurrence sur k . On sait (cf. SANSONE [35] p. 324) qu'il existe une constante c telle que

$$\sup_{j \in \mathbf{N}} \sup_{x \in \mathbf{R}} |h_j(x)| \leq c \quad (14)$$

D'autre part, les relations suivantes entre polynômes d'Hermite sont bien connues :

$$H_{j+1} + 2x H_j + 2j H_{j-1} = 0, \quad \forall j \geq 1$$

$$H'_j = -2j H_{j-1}, \quad \forall j \geq 1$$

$$\text{Or } h'_j(x) = e^{-x^2/2} (2^j j! \pi^{1/2})^{-1/2} [H'_j - x H_j].$$

$$\text{donc } h'_j(x) = e^{-x^2/2} (2^j j! \pi^{1/2})^{-1/2} [-2j H_{j-1} + \frac{H_{j+1}}{2} + j H_{j-1}]$$

on en tire facilement la relation :

$$h'_j(x) = h_{j+1}(x) \frac{\sqrt{j+1}}{\sqrt{2}} - h_{j-1}(x) \frac{\sqrt{j}}{\sqrt{2}} \quad (15)$$

$$\forall j \geq 1, \quad \forall x \in \mathbf{R}$$

Compte tenu de (14), il vient

$$|h'_j(x)| \leq c \left[\frac{\sqrt{j+1}}{\sqrt{2}} + \frac{\sqrt{j}}{\sqrt{2}} \right] < c [2(j+q)]^{1/2} \quad (16)$$

$$\forall x \in \mathbf{R} \quad \forall j \geq 1.$$

Or, de (15), on tire

$$h_j^{(k)}(x) = h_{j+1}^{(k-1)}(x) \frac{\sqrt{j+1}}{\sqrt{2}} - h_{j-1}^{(k-1)}(x) \cdot \frac{\sqrt{j}}{\sqrt{2}} \quad (15 \text{ bis})$$

$$\forall x \in \mathbb{R}, \quad \forall j \geq 1, \quad \forall k \geq 1$$

donc, si l'on suppose que

$$\sup_{x \in \mathbb{R}} |h_j^{(k-1)}(x)| \leq c [2(j+k-1)]^{\frac{k-1}{2}}, \quad \forall j \geq 1$$

on en tire (13) et, compte tenu de (14) et (16), on en déduit que (13) est valable $\forall j \geq 1$ et $\forall k \geq 0$. C.Q.F.D.

LEMME 13 :

$$\forall \alpha > \frac{1}{2}, \quad \lim_{j \rightarrow \infty} \int_{|x| \geq j} |x|^\alpha |h_j(x)| dx = 0$$

DÉMONSTRATION.

On sait ⁽¹⁾ que le $j^{\text{ème}}$ polynôme d'Hermite peut s'écrire :

$$H_j(x) = \sum_{k=0}^{[j/2]} (-1)^k \frac{i(j-1) \dots (j-2k+1)}{k!} (2x)^{j-2k}$$

d'où

$$|H_j(x)| \leq \left(\sum_{k=0}^{[j/2]} \frac{j^k}{k!} \right) (2|x|)^j < (2e|x|)^j$$

donc

$$|h_j(x)| \leq (2^j j! \pi^{1/2})^{-1/2} (2e|x|)^j e^{-\frac{x^2}{2}}$$

d'où, pour j assez grand et compte tenu de la formule de Stirling :

(1) cf SANSONE [35] p. 306.

$$|h_j(x)| \leq K \frac{2^{j/2} e^j}{j^{j/2} e^{-j/2} j^{1/4}} |x|^j e^{-\frac{x^2}{2}}$$

où K est une constante indépendante de j .

Alors, pour j assez grand, il vient :

$$|h_j(x)| \leq |x|^j e^{-\frac{x^2}{2}} \quad (17)$$

mais, pour $|x| \geq j^\alpha$, $\alpha > \frac{1}{2}$, on a

$$j \log |x| \leq |x| \frac{1}{\alpha} \log |x|$$

Ce qui entraîne, pour j assez grand,

$$j \log |x| \leq \frac{x^2}{4} \text{ soit } |x|^j \leq e^{\frac{x^2}{4}}$$

et, en reportant dans (17), on trouve :

$$|h_j(x)| < e^{-\frac{x^2}{4}} \quad (18)$$

d'où l'on déduit immédiatement le résultat cherché.

LEMME 14 :

$\forall \alpha > \frac{1}{2}$, il existe une constante positive K telle que, pour j assez grand

$$\int_{\mathbb{R}} |h_j(x)| \cdot dx \leq K j^{\alpha/2} \quad (19)$$

DÉMONSTRATION.

D'après l'inégalité de Schwarz, on a

$$\int_{|x| \leq j^\alpha} |h_j(x)| \cdot dx \leq \left[\int_{|x| \leq j^\alpha} h_j^2(x) \cdot dx \right]^{1/2} \cdot (2j^\alpha)^{1/2} < (2j^\alpha)^{1/2}$$

Donc, d'après le lemme 13 et l'égalité

$$\int_{\mathbf{R}} |h_j(x)| \, dx = \int_{|x| \leq j^\alpha} |h_j(x)| \, dx + \int_{|x| > j^\alpha} |h_j(x)| \, dx$$

on a, pour j assez grand :

$$\int_{\mathbf{R}} |h_j(x)| \, dx < 2 \cdot j^{\alpha/2} \quad (20)$$

C.Q.F.D.

LEMME 15 :

$\forall \alpha > \frac{1}{2}$ on a, pour j assez grand

$$\frac{1}{2c} \cdot j^{1/2} < \int_{\mathbf{R}} |h'_j(x)| \, dx < 2 \cdot j^{1/2 + \alpha/2} \quad (21)$$

DÉMONSTRATION.

D'après (15), on peut écrire

$$\int_{\mathbf{R}} |h'_j(x)| \, dx \leq \frac{\sqrt{j+1}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j+1}(x)| \, dx + \frac{\sqrt{j}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j-1}(x)| \, dx$$

et en tenant compte de (20), on obtient facilement la deuxième inégalité de (21).

Toujours d'après (15), on peut écrire :

$$h'^{(2)}_j(x) = \frac{j+1}{2} h^2_{j+1}(x) + \frac{j}{2} h^2_{j-1}(x) - \sqrt{j(j+1)} h_{j+1}(x) h_{j-1}(x)$$

ce qui donne, en intégrant :

$$\int_{\mathbf{R}} h'^{(2)}_j(x) \, dx = j + \frac{1}{2}$$

d'où : $j + \frac{1}{2} = \int_{\mathbf{R}} h'^{(2)}_j(x) \, dx < \sup_{x \in \mathbf{R}} |h'_j(x)| \cdot \int_{\mathbf{R}} |h'_j(x)| \, dx$

ce qui entraîne, d'après (13) :

$$\int_{\mathbf{R}} |h'_j(x)| \, dx > \frac{j + \frac{1}{2}}{c [2(j+1)]^{1/2}}$$

et la première inégalité de (21) s'en déduit immédiatement.

LEMME 16 :

$\forall \alpha > \frac{1}{2}$ et $\forall k \geq 1$, on a

$$\int_{\mathbf{R}} |h_j^{(k)}(x)| \, dx = O\left(j^{\frac{k+\alpha}{2}}\right) \quad (22)$$

DÉMONSTRATION.

Nous allons procéder par récurrence.

Le lemme 15 montre que la propriété est vraie pour 1, supposons-la vraie pour $k-1$.

De (15 bis), on tire :

$$\begin{aligned} \int_{\mathbf{R}} |h_j^{(k)}(x)| \, dx &\leq \frac{\sqrt{j+1}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j+1}^{(k-1)}(x)| \, dx \\ &\quad + \frac{\sqrt{j}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j-1}^{(k-1)}(x)| \, dx \end{aligned} \quad (23)$$

Or, d'après l'hypothèse de récurrence :

$$\int_{\mathbf{R}} |h_{j+1}^{(k-1)}(x)| \, dx = O\left(j^{\frac{k-1+\alpha}{2}}\right)$$

et

$$\int_{\mathbf{R}} |h_{j-1}^{(k-1)}(x)| \, dx = O\left(j^{\frac{k-1+\alpha}{2}}\right)$$

en reportant dans (23), on en tire (22). C.Q.F.D.

Nous pouvons maintenant appliquer les théorèmes généraux aux fonctions d'Hermite :

COROLLAIRE 13 :

Si f est $k+3$ fois dérivable si $e^{\frac{x^2}{2}} \cdot [e^{-\frac{x^2}{2}} f(x)]^{(k+3)} \in \mathcal{L}^2$
et s'il existe $\alpha > \frac{1}{2}$ tel que

$$\sum_{n=1}^{\infty} \exp \left[-\gamma \frac{n}{q(n)^{k+3+\alpha}} \right] < +\infty, \quad \forall \gamma > 0 \quad (24)$$

alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \quad \|\hat{f}_n - f\|_k = 0$$

$\hat{f}_n$ désignant l'estimateur associé aux fonctions d'Hermite.

DÉMONSTRATION.

Tout d'abord, d'après les lemmes 11 et 12, on a, $\forall \ell \leq k$:

$$\sup_{x \in \mathbb{R}} |a_j \cdot e_j^{(\ell)}(x)| = O \left(j^{-\frac{k+3}{2}} \cdot j^{\frac{1}{2}} \right)$$

Donc $f \in C_k$.

D'autre part, d'après le lemme 12, on a pour tout $\ell < k$.

$$\mu_{q(n), \ell} \leq \sup_{0 \leq f \leq k} \mu_{q(n), \ell} = O \left((q(n))^{k/2} \right)$$

Enfin, d'après le lemme 15 :

$$V_{q(n)} = O \left((q(n) \cdot q(n))^{\frac{1+\alpha}{2}} \right) = O \left((q(n))^{\frac{3+\alpha}{2}} \right)$$

d'où le résultat d'après le théorème 17.

REMARQUE :

Pour $k = 0$, si l'on suppose seulement f continue à variation bornée, alors sa série de fonctions d'Hermite converge uniformément sur tout compact (cf SANSONE [35] p. 436) et l'on en déduit que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in K} |\hat{f}_n(x) - f(x)| = 0$$

$\forall K$ compact de $\mathbf{R}$, si la condition (24) est vérifiée, ce qui sera le cas, par exemple, pour

$$q(n) = n^\beta, \quad \beta < \frac{2}{7}$$

Donnons maintenant les résultats relatifs à l'estimateur $\hat{f}_{n,k}$ associé aux fonctions d'Hermite :

COROLLAIRE 14 :

Si l'on prend comme système orthonormal les fonctions d'Hermite et si $f \in \mathcal{L}^2(\mathbf{R}, k)$ est telle que

$$\lim_{x \rightarrow -\infty} h_j^{(\ell)}(x) f^{(k-\ell)}(x) \text{ et } \lim_{x \rightarrow +\infty} h_j^{(\ell)}(x) f^{(k-\ell)}(x)$$

existent et soient égales pour $\ell = 0, \dots, k$; alors

$$1^\circ) \text{ Si } \frac{q(n)^{k+1}}{n} \rightarrow 0 \text{ avec } \frac{1}{n}, \hat{f}_{n,k}$$

converge vers $f^{(k)}$ en moyenne quadratique intégrée.

2°) Si f est $k+3$ fois dérivable et telle que

$$e^{x^2/2} \frac{d^3}{dx^3} [e^{-x^2/2} f^{(k)}(x)] \in \mathcal{L}^2$$

et s'il existe $\alpha > \frac{1}{2}$ tel que :

$$\sum_{n=1}^{+\infty} \exp \left[-\gamma \frac{n}{q(n)^{k+\alpha+2}} \right] < +\infty, \quad \forall \gamma > 0 \quad (25)$$

alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in \mathbb{R}} |\hat{f}_{n,k}^{(k)}(x) - f^{(k)}(x)| = 0$$

DÉMONSTRATION.

1°) Le lemme 12 montre que

$$\int_{\mathbb{R}} \frac{1}{n} \left[\sum_{j=1}^{q(n)} [h_j^{(k)}(x)]^2 \right] f(x) dx < \frac{1}{n} c^2 2^k \cdot [q(n) + k]^k \cdot q(n)$$

donc $\frac{q(n)^{k+1}}{n} \rightarrow 0$ quand $n \rightarrow \infty$ entraîne (6) et l'on peut appliquer le 1°) du théorème 15.

2°) Le lemme 16 montre que

$$V_{q(n),k} = O\left(q(n) \cdot q(n)^{\frac{k+\alpha}{2}}\right) = O\left(q(n)^{\frac{k+\alpha+2}{2}}\right)$$

et comme

$$M_{q(n)} < c, \quad \forall n \in \mathbb{N}$$

on voit que (25) entraîne (7) et, compte tenu du lemme 11, on peut appliquer le 2°) du théorème 15.

REMARQUE :

Si l'on suppose seulement $f^{(k)}$ continue et à variation bornée, on obtient, dans le 2°), une convergence uniforme presque sûre sur tout compact.

CONCLUSION.

Pour obtenir une convergence uniforme presque sûre, nous avons dû faire des hypothèses légèrement plus fortes que celles de [14]. Nous pensons que l'on doit pouvoir alléger ces hypothèses mais, de toute façon, elles sont presque toujours vérifiées dans la pratique.

D'autre part, la méthode des fonctions orthogonales a des qualités de simplicité et de robustesse que ne possède pas la méthode du noyau.

Par exemple, pour $n = 1000$, l'estimateur de la densité (ou de l'une de ses dérivées) obtenu par la méthode du noyau se présentera sous la forme d'une somme *irréductible* de mille termes ! En revanche, l'estimateur construit à partir des fonctions d'Hermite se présentera sous la forme du produit d'un polynôme de faible degré par une exponentielle: on obtiendra donc un estimateur à la fois plus simple et plus lisse.

BIBLIOGRAPHIE DES CHAPITRES I, II et III

- [33] BHATTACHARYA (P.K.) (1967). Estimation of a Probability density function and its derivatives.
Sankya, Ser. A 29 p. 373-382.
- [6] BICKEL (P.J.), LEHMANN (E.L.) (1969). Unbiased estimation in convex families.
Ann. Math. Stat., Vol. 40 n° 5 p. 1523-1535.
- [32] BOSQ (D.) (1969). Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales.
Comptes rendus, t. 268 p. 555-557.
- [34] BOSQ (D.) (1969). Estimation non paramétrique de la densité et de ses dérivées.
Comptes rendus, t. 269 p. 1010-1012.
- [48] BOSQ (D.) (1970). Complément à deux notes sur l'estimation de la densité et de ses dérivées.
Comptes rendus, t. 271 p. 45.
- [15] BOURBAKI (1955). Espaces vectoriels topologiques (Chapitre IV).
Hermann, Paris.
- [20] CACOULOS (T) (1966). Estimation of a multivariate density.
Ann. Inst. Stat. Math. 18 p. 179-190.
- [25] CENCOV (N.N.) (1962). Evaluation of an unknown distribution density from observations.
Soviet Math. 3 p. 1559-1562.
- [2] COURREGE (P.) (1965). Théorie de la Mesure.
CDU, Paris.
- [22] CRASWELL (K.J.) (1965). Density estimation in a topological group.
Ann. Math. Stat. 36 p. 1047-1048.
- [9] DUGUE (D.) (1958). Traité de Statistique théorique et appliquée.
Masson, Paris.
- [4] DUNFORD (D.N.), SCHWARTZ (J.T.) (1958). Linear operators (Tome I).
Wiley, New York.

- [13] DVORETZKY (A.), KIEFER (J.). WOLFOWITZ (J.) (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator.
Ann. Math. Stat. 27 p. 642-669.
- [16] FRASER (D.A.S.) (1959). Non parametric Methods in Statistics.
Wiley, New York.
- [5] HALMOS (P.R.) (1946). The theory of unbiased estimation.
Ann. Math. Stat. 17 p. 34-43.
- [29] PARZEN (E.) (1962). On estimation of a probability density function and mode.
Ann. Math. Stat. 33 p. 1065-1076.
- [24] PICKANDS (J.) III (1969). Efficient estimation of a probability density function.
Ann. Math. Stat. 40 p. 854-864.
- [30] RENYI (A.) (1966). Calcul des Probabilités.
Dunod, Paris.
- [7] ROSENBLATT (M.) (1956). Remark on some non parametric estimates of a density function.
Ann. Math. Stat. 27 p. 832-837.
- [35] SANSONE (G.) (1959). Orthogonal functions.
Intersciences, New York.
- [14] SCHUSTER (E.F.) (1969). Estimation of a probability density function and its derivatives.
Ann. Math. Stat. 40 p. 1187-1195.
- [27] SCHWARTZ (S.C.) (1967). Estimation of probability density by an orthogonal series.
Ann. Math. Stat. 38 p. 1261-1265.
- [26] TIAGO DE OLIVEIRA (J.) (1961). Estatística de densidades; Resultados assintóticos.
Rev. Fac. Ci. Univ. Lisboa A 9 1961 p. 165-171.
- [8] TUKEY (J.W.) (1961). Curves as parameters, and touch estimation.
Proc. 4 th. Symp. Math. Stat. Prob.,
Berkeley, p. 681-694.

- [10] JILOVEC (S.) (1967). On the existence of superconsistent estimates.
4th. conf. Inf. th. Stat. Dec. funct. Rand. Proc. Prague 1965 Trans.
Prague, Academia, 24 p. 343-380.
- [28] KRONMAL (R.) et TARTER (M.) (1968). The estimation of probability
densities and cumulatives by Fourier series methods.
J. Amer. Stat. Assoc., 63 p. 925-952.
- [17] LANNUZEL (B) (1969). Thèse 3ème cycle.
Paris.
- [23] LEADBETTER (M.R.), WATSON (G.S.) (1963). On the estimation of the
probability density I.
Ann. Math. Stat. 34 p. 480-491.
- [49] LE CAM (L.) (1953). On some asymptotic properties of maximum likeli-
hood and related Bayes estimates.
Univ. Calif. Publ. Stat. Vol. 1 p. 277-300.
- [12] LE CAM (L.) (1956). On the asymptotic theory of estimation and testing
hypotheses.
Proc. of 3^d Berkeley symposium p. 129-156.
- [11] LE CAM (L.), SCHWARTZ LORRAINE (1960). A necessary and suffi-
cient condition for the existence of consistent estimates.
Ann. Math. Stat. 31 p. 140-150.
- [3] LEHMANN (E.L.) (1951). A general concept of unbiasedness.
Ann. Math. Stat. 22 p. 587-597.
- [31] LOEVE (M.) (1955). Probability Theory.
Van Nostrand, New York.
- [18] NADARAYA (E.A.) (1965). On non-parametric estimates of density
functions and regression curves.
Theor. Prob. Appl. 10 p. 186-190.
- [1] PARZEN (E.) (1958). On asymptotically efficient consistent estimates
of the spectral density of a stationary time series.
J. Roy. Stat. Soc. Ser B 20 p. 303-322.

- [21] VAN RYZIN (J.) (1969). On strong consistency of density estimates.
Ann. Math. Stat. 40 p. 1765-1772.

- [19] WOODROOFE (M.) (1967). On the maximal deviation of the sample density.
Ann. Math. Stat. 38 p. 475-481.

CONTRIBUTION A LA THEORIE DE

L'ESTIMATION FONCTIONNELLE

D. BOSQ

TABLE DES MATIÈRES

	Pages
Chapitre IV - ESTIMATION DE LA DENSITÉ CONDITIONNELLE ET DES FONCTIONS DE RÉGRESSION	99
I - Généralités et notations	99
II - Estimation de la densité conditionnelle	100
III - Estimation des fonctions de régression	115
Chapitre V - ESTIMATION DE LA FONCTION FRONTIÈRE	122
I - Position du Problème	122
II - Construction d'un estimateur convergent	123
III - Cas d'un espace localement compact	128
IV - Cas où la fonction frontière est de signe quelconque	131
V - Application: Définition d'une méthode de Monte Carlo	140
Chapitre VI - ÉLÉMENTS EXTRÉMAUX D'UN ÉCHANTILLON VECTORIEL	142
I - Généralités	142
II - Loi du nombre d'éléments maximaux associés à une rotation de $\mathbf{R}^2$	145
III - Comportement asymptotique du nombre d'éléments maximaux (resp. minimaux)	150
IV - Théorèmes de M. Geffroy - Application à l'estimation de la courbe maximale	166
BIBLIOGRAPHIE	173

IV

ESTIMATION DE LA DENSITÉ CONDITIONNELLE ET DES FONCTIONS DE RÉGRESSION

I - GÉNÉRALITÉS ET NOTATIONS

Dans ce chapitre nous montrons qu'à partir de toute suite d'estimateurs de la densité possédant certaines propriétés asymptotiques, on peut construire une suite d'estimateurs de la régression qui converge uniformément presque sûrement. Pour cela, nous passons par l'intermédiaire d'un estimateur de la densité conditionnelle.

Nous connaissons un seul article sur cette question ⁽¹⁾: celui de BHATTACHARYA et PARTHASARATHY [36] (1961). Ces deux auteurs utilisent la méthode de l'analyse graphique fractile.

Les résultats que nous donnons sont ceux de notre note aux C.R. du 13 Octobre 1969.

Les notations que nous emploierons sont les suivantes :

- $F(x,y)$ est la fonction de répartition du couple de variables aléatoires (X, Y) . Nous supposons que F est absolument continue, que sa densité $f(x,y)$ est bornée et qu'elle est définie partout par la relation

$$\frac{\partial^2 F}{\partial x \partial y} (x,y) = f(x,y).$$

- $g(x)$ désigne la densité marginale de X .

- $h_{x,a}(y)$ est la densité de Y conditionnée par

$$X \in [x-a, x+a], \quad x \in \mathbb{R}, \quad a > 0.$$

- $h_x(y)$ est la densité de Y conditionnée par $X = x, x \in \mathbb{R}$.

- $B(x,a) = \{(x', y') \in \mathbb{R}^2 \mid x-a \leq x' < x+a, y' \in \mathbb{R}\}$

(1) A part celui de NADARAYA [18] qui traite un cas particulier.

- $p(x,a) = \text{Prob} [(X, Y) \in B(x,a)]$

- $(x_i, y_i)_{1 \leq i \leq n}$ est un échantillon de taille n tiré suivant F .

- $l_n(x,a)$ est le nombre de points de l'échantillon appartenant à $B(x,a)$; $l_n(x,a)$ est la réalisation d'une variable aléatoire $L_n(x,a)$.

Enfin, $y_{i_1}, \dots, y_{i_{l_n}}$ sont les ordonnées des points de l'échantillon appartenant à $B(x,a)$ ⁽¹⁾.

Nous supposons que a est une fonction positive de

$n(a = a(n) > 0, \forall n \in \mathbb{N})$ et telle que

$$\lim_{n \rightarrow \infty} a(n) = 0 \quad (1)$$

II - ESTIMATION DE LA DENSITÉ CONDITIONNELLE

1. Définition d'une classe d'estimateurs.

Pour définir un estimateur de $h_X(y)$, il semble naturel de partir d'un estimateur de la densité univariée.

Comme nous n'aurons besoin, dans la suite, que de certaines propriétés d'un estimateur de la densité, il est inutile de préciser la méthode utilisée pour construire cet estimateur.

Nous désignons donc par $\hat{f} = (\hat{f}_n)_{n=0,1,2,\dots}$ une suite d'estimateurs de la densité univariée ⁽²⁾. Pour des raisons qui apparaîtront ci-dessous, nous posons arbitrairement :

$$\hat{f}_0(y) = 0, \quad \forall y \in \mathbb{R} \quad (2)$$

Soit alors $\mathcal{F}(\hat{f})$ l'ensemble des densités univariées telles que :

$$1^\circ) f_n^g(y) = E_g[(\hat{f}_n(y))] = \int_{\mathbb{R}^n} \hat{f}_n(y_1, \dots, y_n; y) g(y_1) \dots g(y_n) dy_1 \dots dy_n$$

(1) Pour abrégé, nous écrirons l_n au lieu de $l_n(x,a)$ quand il n'y aura pas ambiguïté.

(2) Nous supposons, une fois pour toutes, que $\hat{f}_n(y_1, \dots, y_n; y)$ est une application mesurable de $\mathbb{R}^{n+1}$ dans $\mathbb{R}$, $\forall n \in \mathbb{N}$.

existe $\forall y \in \mathbf{R}, \forall n \in \mathbf{N}, \forall g \in \mathcal{F}(\hat{f})$.

$$2^{\circ}) \sup_{y \in \mathbf{R}} |f_n^g(y) - g(y)| \rightarrow 0 \quad \text{avec} \quad \frac{1}{n}, \quad \forall g \in \mathcal{F}(\hat{f}).$$

3°) Il existe deux constantes positives C_1 et C_2 et une fonction de n , α_n , non décroissante, tendant vers $+\infty$ avec n et vérifiant la condition

$$\lim_{n \rightarrow \infty} \frac{n}{\alpha_n^2} = +\infty, \quad \text{telles que}$$

$$\text{Prob} \left[\sup_{y \in \mathbf{R}} |\hat{f}_n(y) - f_n^g(y)| > \varepsilon \right] \leq C_1 \exp \left[-C_2 \frac{n}{\alpha_n^2} \varepsilon^2 \right]$$

$\forall \varepsilon > 0, \forall n \in \mathbf{N}, \forall g \in \mathcal{F}(\hat{f})$.

REMARQUE :

Les conditions précédentes impliquent que

$$\sup_{y \in \mathbf{R}} |\hat{f}_n(y) - g(y)| \xrightarrow{p} 0 \quad \text{quand} \quad n \rightarrow +\infty$$

$\forall g \in \mathcal{F}(\hat{f})$.

$\hat{f}$ étant donnée et x étant fixé, nous prendrons comme estimateur de h_x la fonction aléatoire définie par :

$$\hat{h}_{n,x}(y) = \hat{f}_{1_n}(y_{1_1}, \dots, y_{1_{1_n}}; y) \quad (3)$$

$\forall y \in \mathbf{R}, \forall n \in \mathbf{N}$.

2. Convergence de l'estimateur.

Dans ce paragraphe, x est fixé une fois pour toutes.

LEMME 17 :

Si $g(x) > 0$ et s'il existe $\delta > 0$ tel que, pour $a \leq \delta$, f est

lipchitzienne d'ordre γ_1 dans $B(x, a)$ et g est lipchitzienne d'ordre γ_2 dans $[x-a, x+a]$, alors, il existe $\delta_1 \in]0, \delta]$ tel que $a < \delta_1$ entraîne

$$\sup_{y \in \mathbb{R}} |h_{x,a}(y) - h_x(y)| \leq A(x) \cdot a^\gamma \quad (4)$$

où $A(x)$ est indépendante de y et de a et où l'on a posé

$$\gamma = \max(\gamma_1, \gamma_2).$$

DÉMONSTRATION.

Si $g(x) > 0$, on a $F(x+a, +\infty) - F(x-a, +\infty) > 0$, $\forall a > 0$ et l'on peut écrire :

$$h_{x,a}(y) = \frac{\frac{\partial F}{\partial y}(x+a, y) - \frac{\partial F}{\partial y}(x-a, y)}{F(x+a, +\infty) - F(x-a, +\infty)}$$

Or, pour $a \leq \delta$, f est Lipchitzienne donc continue et, a fortiori, continue par rapport à x ; il en est de même pour g et l'on peut écrire :

$$h_{x,a}(y) = \frac{f(x-a + 2\theta a, y)}{g(x-a + 2\theta' a)}, \quad 0 < \theta < 1, \quad 0 < \theta' < 1$$

D'autre part, $h_x(y) = \frac{f(x, y)}{g(x)}$ et comme, pour a assez petit ($a < \delta_1$) on

a $g(x-a + 2\theta' a) > \frac{g(x)}{2}$, il vient :

$$\begin{aligned} |h_{x,a}(y) - h_x(y)| &\leq \frac{2}{g(x)} |f(x-a + 2\theta a, y) - f(x, y)| \\ &\quad + \frac{2f(x, y)}{g(x)^2} |g(x) - g(x-a + 2\theta' a)| \end{aligned}$$

Or, il existe deux constantes k_1 et k_2 telles que, $\forall a \leq \delta$

$$|f(x-a + 2\theta a, y) - f(x, y)| \leq k_1 a^{\gamma_1} |1 - 2\theta|^{\gamma_1} \leq k_1 a^{\gamma_1}$$

et

$$|g(x) - g(x-a + 2\theta' a)| \leq k_2 a^{\gamma_2} |1 - 2\theta'|^{\gamma_2} \leq k_2 a^{\gamma_2}$$

d'où :

$$|h_{x,a}(y) - h_x(y)| \leq \frac{2}{g(x)} k_1 a^{\gamma_1} + \frac{2f(x,y)}{g(x)^2} k_2 a^{\gamma_2}$$

Or, on sait que $M = \sup_{(x,y) \in \mathbf{R}^2} f(x,y) < +\infty$, on a donc immédiatement

l'inégalité (4) avec

$$A(x) = \frac{2}{g(x)} k_1 + \frac{2M}{g(x)^2} k_2 \quad (5)$$

Pour simplifier l'énoncé du lemme suivant, nous allons poser

$$h_{n,x}(y) = f_{L_n(x,a)}^{h_{x,a}}(y) \quad \forall n \in \mathbf{N}, \quad \forall y \in \mathbf{R} \quad (6)$$

Autrement dit, $\forall y \in \mathbf{R}$ et $\forall n \in \mathbf{N}$, $h_{n,x}(y)$ est une variable aléatoire liée à $L_n(x,a)$ par la relation fonctionnelle suivante :

$$L_n(x,a) = \ell \implies h_{n,x}(y) = E_{h_{x,a}}[\hat{f}_1(y)] \quad (7)$$

$\forall \ell = 0, 1, \dots, n$.

Nous avons alors le

LEMME 18 :

Sous les hypothèses du lemme 17 et en supposant, de plus, que $h_{x,a} \in \mathcal{F}(\hat{f})$ au moins pour a assez petit et que

$$\lim_{n \rightarrow \infty} \frac{n a_n}{\gamma_n^2} = +\infty \quad \text{on a}$$

$$\sup_{y \in \mathbf{R}} | \hat{h}_{n,x}(y) - h_{n,x}(x) | \xrightarrow{p} 0 \quad \text{quand } n \rightarrow \infty. \quad (8)$$

DÉMONSTRATION.

Désignons par d la distance de la convergence uniforme. En appliquant la formule des probabilités totales, on obtient :

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon] = \sum_{\ell=0}^n C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell}$$

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon / L_n = \ell]$$

Or, d'après (7), on a

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon / L_n = \ell] = \text{Prob} [d(\hat{f}_1; f_\ell^{h_{x,a}}) > \varepsilon]$$

cette dernière probabilité étant calculée par rapport à la loi conditionnelle de densité $h_{x,a}$.

Comme on a supposé que, pour a assez petit, $h_{x,a} \in \mathcal{F}(\hat{f})$, on peut, pour n assez grand, appliquer 3°) et écrire :

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon] \leq C_\ell \sum_{\ell=0}^n C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell}$$

$$\exp \left[-C_2 \frac{1}{\alpha_\ell^2} \varepsilon^2 \right]$$

et comme α_1 a été supposée non décroissante :

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon] \leq C_\ell \sum_{\ell=0}^n C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell}$$

$$\exp \left[-C_2 \frac{1}{\alpha_n^2} \varepsilon^2 \right]$$

$$\leq C_1 [1-p_{x,a} [1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})]]^n$$

Or, pour a assez petit

$$\begin{aligned}
 p(x,a) &= F(x+a, +\infty) - F(x-a, +\infty) \\
 &= 2a g(x) \quad , \quad x \in]x-a, x+a[\\
 &> 2a \cdot \frac{g(x)}{2} = a \cdot g(x)
 \end{aligned}$$

il vient alors :

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon] \leq C_1 [1 - a_n g(x) [1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})]^n] \quad (9)$$

Le deuxième membre de cette inégalité tend vers 0 si et seulement si :

$$n \text{ Log} [1 - a_n g(x) [1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})]] \rightarrow -\infty$$

donc, compte tenu du fait que $\lim_{n \rightarrow \infty} a_n = 0$ si

$$n a_n [1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})] \rightarrow +\infty$$

or $\alpha_n \rightarrow +\infty$ donc $1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2}) \sim \frac{C_2 \varepsilon^2}{\alpha_n^2}$ et la condi-

tion s'écrit finalement

$$\lim_{n \rightarrow \infty} \frac{n a_n}{\alpha_n^2} = +\infty \quad (10)$$

C.Q.F.D.

LEMME 18 bis :

Sous les hypothèses du lemme 18 et l'hypothèse supplémentaire

$$\sum_{n=0}^{+\infty} \exp[-\lambda \frac{n a_n}{\alpha_n^2}] < +\infty, \quad \forall \lambda > 0 \quad (11)$$

On a $d(\hat{h}_{n,x}; h_{n,x}) \xrightarrow{\text{p.s.}} 0$ avec $\frac{1}{n}$.

DÉMONSTRATION.

On sait, puisque $\text{Log}(1-u) \leq -u$, $\forall u < 1$, que

$$n \text{Log} \left[1 - a_n g(x) \right] 1 - \exp \left(-C_2 \frac{\varepsilon^2}{\alpha_n^2} \right) \leq -n a_n g(x) \left[1 - \exp \left(-C_2 \frac{\varepsilon^2}{\alpha_n^2} \right) \right]$$

pour n assez grand.

(12)

D'autre part, on a, pour n assez grand :

$$1 - \exp \left(-\frac{C_2 \varepsilon^2}{\alpha_n^2} \right) < 2 \frac{C_2 \varepsilon^2}{\alpha_n^2} \quad (13)$$

(12) et (13) entraînent alors :

$$n \text{Log} \left[1 - a_n g(x) \right] 1 - \exp \left(-\frac{C_2 \varepsilon^2}{\alpha_n^2} \right) \leq -2 C_2 \varepsilon^2 g(x) \cdot \frac{n a_n}{\alpha_n^2}$$

D'où, d'après (9) et toujours pour n assez grand :

$$\boxed{\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon] \leq C_1 \exp \left[-2 C_2 \varepsilon^2 g(x) \frac{n a_n}{\alpha_n^2} \right]} \quad (14)$$

La condition (11) permet alors de conclure.

Pour le lemme suivant, nous aurons besoin d'une nouvelle propriété de la classe d'estimateurs utilisée: nous dirons que $\hat{f}$ est *uniformément asymptotiquement sans biais* sur $\mathcal{F}'(\hat{f}) \subset \mathcal{F}(\hat{f})$ si la condition 2°) est valable uniformément pour $g \in \mathcal{F}'(\hat{f})$, autrement dit si

$$\lim_{n \rightarrow \infty} \sup_{g \in \mathcal{F}'(\hat{f})} d(f_n^g; g) = 0 \quad (15)$$

Nous dirons, pour abréger, que $\hat{f}$ est u.a.b. sur $\mathcal{F}'(\hat{f})$.

LEMME 19 :

Si $g(x) > 0$, si g est Lipchitzienne dans un voisinage de x , si $h_{x,a} \in \mathcal{F}(\hat{f})$, $\forall n \in \mathbb{N}$, si $\hat{f}$ est u.a.b. sur $\{h_{x,a_n}\}_{n \in \mathbb{N}}$

et si $n a_n \rightarrow +\infty$

alors

$$d(h_{n,x}; h_{x,a}) \xrightarrow{p} 0.$$

DÉMONSTRATION.

D'après (6), on peut écrire :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] = \sum_{\ell=0}^n C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell}$$

$$\text{Prob} [d(f_{\ell}^{h_{x,a}}; h_{x,a}) > \varepsilon / L_n = \ell] \quad (16)$$

Les probabilités qui figurent au 2ème membre de (16) valant 0 ou 1 puisque $d(f_{\ell}^{h_{x,a}}; h_{x,a})$ est une variable certaine quand $L_n = \ell$. Comme $\hat{f}$ est u.a.b. sur $\{h_{x,a_n}\}_{n \in \mathbb{N}}$, il existe un entier ℓ_0 indépendant de a_n (donc de n) tel que $\ell > \ell_0$ entraîne $d(f_{\ell}^{h_{x,a}}; h_{x,a}) < \varepsilon$, $\forall a_n$ donc $\forall n \in \mathbb{N}$.

Dans ces conditions, pour $n > \ell_0$, on peut écrire :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] \leq \sum_{\ell=0}^{\ell_0} C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell} \quad (17)$$

Or, pour n assez grand, on aura

$$\ell_0 < n p(x,a) - [1-p(x,a)] \quad (18)$$

En effet, on a vu dans la démonstration du lemme 18 que, à partir d'un certain rang, on a l'inégalité $p(x,a) > a g(x)$; et comme, par hypothèse, $n a_n \rightarrow +\infty$, on aura toujours $\ell_0 < n a_n g(x)$, à partir d'un certain rang, ce qui entraîne (18).

Donc, pour $n \geq v$, $C_n^\ell p(x,a)^\ell [1-p(x,a)]^{n-\ell}$ sera fonction non

décroissante de ℓ pour $\ell \leq \ell_0$ (propriété connue de la loi binominale).

Dans ces conditions, (17) implique :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] \leq \ell_0 C_n^{\ell_0} p(x,a)^{\ell_0} [1-p(x,a)]^{n-\ell_0} \quad (19)$$

et en choisissant n assez grand pour que

$$a g(x) < p(x,a) < 3a g(x) \quad (19 \text{ bis})$$

on obtient facilement :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] \leq \ell_0 \cdot \frac{n^{\ell_0}}{\ell_0!} [3a g(x)]^{\ell_0} \exp [-(n-\ell_0) a g(x)] \quad (20)$$

A partir de (20), on obtient donc une inégalité de la forme :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] \leq A (n a_n)^{\ell_0} e^{-B n a_n} \quad (21)$$

et, à partir d'un certain rang

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon] \leq A e^{-B n a_n} \quad (22)$$

où A et B sont indépendants de n .

L'hypothèse $n a_n \rightarrow +\infty$ permet de conclure.

LEMME 19 bis :

Si les conditions du lemme 19 sont vérifiées et si, de plus,

$$\sum_{n=0}^{+\infty} \exp (-\gamma n a_n) < +\infty, \quad \forall \gamma > 0 \quad (23)$$

alors : $d(h_{n,x}; h_{x,a}) \xrightarrow{\text{p.s.}} 0$.

DÉMONSTRATION.

C'est une conséquence immédiate de l'inégalité (22).

THÉORÈME 19 :

Sous les conditions suivantes :

1°) Pour a assez petit, f est Lipchitzienne dans $B(x, a)$ et g est Lipchitzienne dans $[x-a, x+a]$.

2°) $h_{x, a_n} \in \mathcal{F}(\hat{f})$, $\forall n \in \mathbf{N}$ et $\hat{f}$ est u.a.b. sur $\{h_{x, a_n}\}_{n \in \mathbf{N}}$.

3°) $\sum_{n=0}^{+\infty} \exp \left[-\lambda \frac{n a_n}{\alpha_n^2} \right] < +\infty$, $\forall \lambda > 0$

on a

$$d(\hat{h}_{n,x}; h_x) \xrightarrow{\text{p.s.}} 0.$$

DÉMONSTRATION.

On écrit :

$$d(\hat{h}_{n,x}; h_x) \leq d(\hat{h}_{n,x}; h_{n,x}) + d(h_{n,x}; h_{x,a}) + d(h_{x,a}; h_x) \quad (24)$$

et il suffit d'appliquer les lemmes 17, 18 bis et 19 bis pour conclure.

3. Etude de la vitesse de convergence.

Les inégalités (4), (14) et (22) montrent que, pour n assez grand, on a l'inégalité :

$$\text{Prob} \left[d(\hat{h}_{n,x}; h_x) > \varepsilon \right] \leq K_1 \exp \left[-\frac{n a_n}{\alpha_n^2} K_2 \varepsilon^2 \right] \quad (25)$$

En examinant les démonstrations des lemmes précédents, on constate que (25) reste valable, uniformément sur un intervalle compact, si les hypothèses faites sont uniformes sur cet intervalle (cf. les hypothèses du lemme 17, la formule (5), les inégalités (19 bis), la formule (15), l'inégalité (14) et les valeurs de A et B dans l'inégalité (22). On peut donc énoncer le

THÉORÈME 20 :

Soit $[u, v]$ un intervalle compact tel que :

1°) Il existe un voisinage V de $[u, v]$ dans $\mathbf{R}$ tel que f soit Lipchitzienne sur $V \times \mathbf{R}$ et g Lipchitzienne sur V .

2°) $g(x) > 0$, $\forall x \in [u, v]$.

3°) $h_{x, a_n} \in \mathcal{F}(\hat{f})$, $\forall x \in [u, v]$, $\forall n \in \mathbf{N}$.

4°) $\hat{f}$ est u.a.b. sur $\{h_{x, a_n}\}_{u \leq x \leq v}$; $n \geq v$ ($v \in \mathbf{N}$)

Alors si $\frac{n a_n}{\alpha_n^2} \rightarrow \infty$ et si $\varepsilon > 0$ est donné, il existe deux nombres positifs $K_1(\varepsilon)$ et $K_2(\varepsilon)$ et un entier $N(\varepsilon)$ tels que

$$\sup_{u \leq x \leq v} \text{Prob} [d(\hat{h}_{n,x}; h_x) > \varepsilon] \leq K_1(\varepsilon) \exp \left[-K_2(\varepsilon) \frac{n a_n}{\alpha_n^2} \right] \quad (26)$$

$$\forall n \geq N(\varepsilon).$$

Pour préciser un peu plus la vitesse de convergence, supposons que l'on ait :

$$\beta_n = \sup_{\substack{u \leq x \leq v \\ k \geq v}} d(f_n^{h_{x, a_k}}; h_{x, a_k}) \leq C'_1 n^{-b}, \quad C'_1 > 0, b > 0 \quad (27)$$

$$\alpha_n = C'_2 \cdot n^c, \quad 0 < c < \frac{1}{2} \quad (28)$$

$$a_n = C'_3 \cdot n^{-d}, \quad 0 < d < 1 - 2c \quad (29)$$

On peut énoncer alors le

COROLLAIRE 15 :

Si $\gamma_1 = \gamma_2 = 1$ et si $0 < \delta < \inf(d, b(1-d), \frac{1-d}{2} - c)$

Alors, sous les hypothèses du théorème 20 :

$$n^{\delta} \cdot d(\hat{h}_{n,x}; h_x) \xrightarrow[p.s.]{} 0, \quad \forall x \in [u, v].$$

DÉMONSTRATION.

Tout d'abord, compte tenu de (4), (29) et du fait que $\delta < d$, on a $\forall x \in [u, v]$:

$$n^{\delta} \cdot d(h_{x,a}; h_x) \rightarrow 0 \quad (30)$$

D'autre part, si l'on reprend les démonstrations des lemmes 18 et 18 bis, on arrive à l'inégalité :

$$\text{Prob} [d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon \cdot n^{-\delta}] \leq C_1 \cdot \exp [-C_2 \varepsilon^2 n^{-2\delta} g(x) \frac{n a_n}{\alpha_n^2}] \quad (14 \text{ bis})$$

et le deuxième membre de l'inégalité (14 bis) est le terme général d'une série convergente puisque

$$n^{-2\delta} \cdot \frac{n a_n}{\alpha_n^2} = \frac{C'_3}{C'^2_2} \cdot n^{1-d-2\delta-2c}$$

et que $\delta < \frac{1-d}{2} - c$. On a donc, $\forall x \in [u, v]$:

$$n^{\delta} \cdot d(\hat{h}_{n,x}; h_{n,x}) \xrightarrow[p.s.]{} 0 \quad (31)$$

Enfin, si l'on reprend la démonstration du lemme 19, on constate que l'inégalité :

$$d(f_{\ell}^{h_{x,a}}; h_{x,a}) < \varepsilon \cdot n^{-\delta}$$

est vérifiée pour $\ell > \ell_0$ avec ℓ_0 tel que

$$C'_\ell \cdot \ell_0^{-b} < \varepsilon \cdot n^{-\delta} \quad (32)$$

En prenant $\ell_0 = \ell_0(n, \varepsilon) = K n^{\frac{\delta}{b}}$ avec $K > (\frac{C'_\ell}{\varepsilon})^{\frac{\ell}{b}}$, on a (32) et comme

$\delta < b$, on a aussi $\ell_0 < n$, au moins pour n assez grand: les inégalités (18), (19) et (20) restent donc valables si l'on remplace ε par $\varepsilon n^{-\delta}$. Il en est de même pour (21) puisque :

$$\exp [a_n \ell_0 g(x)] \rightarrow \ell \quad \text{quand } n \rightarrow +\infty.$$

D'autre part, comme

$$\text{Log } (n a_n)^{\ell_0} = K n^{\frac{\delta}{b}} (1-d) \text{Log } C'_3 \ell^{-d} . n$$

On a, $\forall u > 0$

$$(n a_n)^{\ell_0} < e^{n^{\frac{\delta}{b} + u}}$$

au moins pour n assez grand. En choisissant $u < 1-d\frac{\delta}{b}$, il vient :

$$\text{Prob} [d(h_{n,x}; h_{x,a}) > \varepsilon n^{-\delta}] < A' e^{-B' n^{\alpha_n}}$$

où A' et B' sont positifs et indépendants de n . Donc :

$$n^{\delta} d(h_{n,x}; h_{x,a}) \xrightarrow{\text{p.s.}} 0 \quad (33)$$

ce qui, joint à (30) et (31), permet de conclure.

4. Exemples.

Examinons comment on peut appliquer les résultats obtenus aux deux classes usuelles d'estimateurs de la densité.

a) *Méthode du noyau.*

L'estimateur utilisé est de la forme :

$$\hat{f}_n(y) = \frac{\alpha_n}{n} \sum_{i=1}^n K[\alpha_n(y-y_i)] \quad (34)$$

où les y_i sont les points de l'échantillon et où K est une densité à variation bornée. Alors d'après un théorème de NADARAYA [18] on sait que

si g est une densité uniformément continue, $g \in \mathcal{F}(\hat{f})^{(1)}$.

De plus, $\forall g \in \mathcal{F}(\hat{f})$, on a l'inégalité : ⁽²⁾

$$d(E_g(\hat{f}_n), g) \leq \sup_{y \in \mathbf{R}} \sup_{|z| \leq \delta} |g(y-z) - g(y)| \\ + 2 \sup_{y \in \mathbf{R}} g(y) \cdot \int_{|u| \geq \delta \alpha_n} K(u) du$$

$\forall \delta > 0, \quad \forall n \in \mathbf{N}.$

On en déduit immédiatement que si $(g_i)_{i \in I}$ est une famille de densités uniformément continues, uniformément bornées et uniformément équicontinues, alors $\hat{f}$ est u.a.b. sur $(g_i)_{i \in I}$.

CONSEQUENCE.

Pour $\hat{f}_n$, si les hypothèses 1°) et 2°) du théorème 20 sont vérifiées, les hypothèses 3°) et 4°) le sont aussi.

$$\text{Car la relation } h_{x,a}(y) = \frac{f(x-a+2\theta a, y)}{g(x-a+2\theta' a)} \quad (35)$$

$$0 < \theta < 1, \quad 0 < \theta' < 1$$

montre que la famille $\{h_{x,a_n}\}_{n \in \mathbf{N}}$ vérifie les conditions indiquées ci-dessus.

Enfin, d'après les résultats de SCHUSTER ⁽³⁾, on voit que, sous les hypothèses 1°) et 2°) du théorème 20, et en supposant, de plus.

$$\sup_{n \in \mathbf{N}} |h'_{x,a_n}| = D < +\infty, \quad \text{si l'on choisit } \delta \in \left] 0, \frac{1}{4} \right[\quad \text{on aura}$$

$$n^\delta \cdot d(\hat{h}_{n,x}; h_x) \xrightarrow{\text{p.s.}} 0, \quad \forall x \in [u, v].$$

(1) La réciproque est vraie moyennant des hypothèses supplémentaires sur K (cf SCHUSTER [14]).

(2) cf [18] p.

(3) Lemme 2.3 et Corollaire 2.6.

b) Méthode des fonctions orthogonales.

On utilise l'estimateur défini au chapitre II en prenant, par exemple, les fonctions d'Hermite comme système orthonormal et $q(n)$ tel que

$$\sum_{n=1}^{\infty} \exp \left[-\gamma \frac{n}{q(n)^4} \right] < +\infty, \quad \forall \gamma > 0.$$

D'après l'inégalité (23 bis) (Chapitre II) et les lemmes 11 et 15 (Chapitre III), on voit que $\alpha_n = q(n)^2$ et que si g est une densité telle que

$$e^{y^2/2} \frac{d^3}{dy^3} [e^{-y^2/2} g(y)]$$

existe et soit de carré intégrable, alors $g \in \mathcal{F}(\hat{f})$.

D'autre part, si $(g_i)_{i \in I}$ est une famille de densités telles que

$$A = \sup_{i \in I} \left\| e^{y^2/2} \frac{d^3}{dy^3} [e^{-y^2/2} g_i(y)] \right\|_{L^2} < +\infty$$

alors $\hat{f}$ est u.a.b. sur $(g_i)_{i \in I}$. On aura, en effet, une majoration de la forme ⁽¹⁾.

$$\sup_{i \in I} d(E_{g_i}(\hat{f}_n); g_i) \leq \frac{A \cdot c}{2^{3/2}} \cdot \sum_{j > q(n)} \frac{1}{j^{3/2}}$$

CONSÉQUENCE.

Si $f(x,y)$ admet une dérivée partielle troisième $f'''_{y^3}(x,y)$ et s'il existe $\eta > 0$ tel que

$$\sup_{|x-x_0| \leq \eta} \left\| e^{y^2/2} \frac{d^3}{dy^3} [e^{-y^2/2} f(x,y)] \right\|_{L^2} < +\infty$$

On en déduit, d'après (35), que $\hat{f}$ est u.a.b. sur $\{h_{x_0, a_n}\}_{n \geq \nu}$
 ν étant convenablement choisi.

(1) cf S.C. SCHWARTZ [27] lemme 3 p. 1262.

III - ESTIMATION DES FONCTIONS DE RÉGRESSION.

1. Cas où Y est bornée.

Soit $[u, v]$ un intervalle compact pour lequel les hypothèses du théorème 20 sont vérifiées. On suppose que $|Y| \leq M$ où M est un nombre positif connu.

Choisissons a_n tel que $k_n = \frac{v-u}{2a_n}$ soit entier, $\forall n \in \mathbb{N}$ et posons :

$$u_1 = u + a_n, u_2 = u + 3a_n, \dots, u_{k_n} = u + (2k_n - 1)a_n.$$

Soit alors $E(x) = \int_{-M}^{+M} y^\alpha h_x(y) dy$, $\alpha > 0$, une fonction de régression de Y en X . On définira un estimateur de $E_\alpha(x)$ en posant :

$$\hat{E}_{n,\alpha}(x) = \int_{-M}^{+M} y^\alpha \hat{h}_{n,u_i}(y) dy \quad \text{pour } x \in [u + (2i-2)a_n, u + 2ia_n[$$

$$i = 1, \dots, k_n^{-1}$$

et

$$\hat{E}_{n,\alpha}(x) = \int_{-M}^{+M} y^\alpha \hat{h}_{n,u_{k_n}}(y) dy \quad \text{pour } x \in [u + (2k_n-2)a_n, u + 2k_na_n]$$

THÉORÈME 21 :

Sous les hypothèses du théorème 20 et l'hypothèse supplémentaire

$$\sum_{n=0}^{\infty} a_n^{-1} \exp \left[-\lambda \frac{n a_n}{\alpha_n^2} \right] < +\infty, \quad \forall \lambda > 0 \quad (36)$$

$$\text{on a } \lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in [u, v]} |\hat{E}_{n,\alpha}(x) - E_\alpha(x)| = 0, \quad \forall \alpha > 0.$$

DÉMONSTRATION.

Tout d'abord, compte tenu des hypothèses 1°) et 2°) du théorème

20, on voit immédiatement que $h_x(y)$ est Lipchitzienne sur $[u, v] \times \mathbf{R}$, d'où l'on déduit que $E_\alpha(x)$ est Lipchitzienne sur $[u, v]$, $\forall \alpha > 0$. Il existe donc deux constantes positives δ et γ telles que

$$|E_\alpha(x) - E_\alpha(u_i)| \leq \delta |x - u_i|^\gamma, \quad \forall x \in [u, v], \quad \forall i \in \{1, \dots, k_n\}$$

d'où, pour $x \in [u_i - a_n, u_i + a_n]$

$$|E_\alpha(x) - E_\alpha(u_i)| \leq \delta \cdot a_n^\gamma, \quad \forall i \in \{1, \dots, k_n\}$$

Compte tenu du fait que $\lim_{n \rightarrow \infty} a_n = 0$ et de l'inégalité

$$|\hat{E}_{n,\alpha}(x) - E_\alpha(x)| \leq |\hat{E}_{n,\alpha}(u_i) - E_\alpha(u_i)| + |E_\alpha(u_i) - E_\alpha(x)|$$

$\forall x \in [u_i - a_n, u_i + a_n], \quad \forall i \in \{1, \dots, k_n\}$, il vient pour n assez grand

$$\begin{aligned} \text{Prob} \left[\sup_{x \in [u, v]} |\hat{E}_{n,\alpha}(x) - E_\alpha(x)| > \varepsilon \right] &\leq \text{Prob} \left[\sup_{1 \leq i \leq k_n} |\hat{E}_{n,\alpha}(u_i) - E_\alpha(u_i)| \right. \\ &\quad \left. > \frac{\varepsilon}{2} \right] \end{aligned} \quad (37)$$

où ε est un nombre positif donné.

Or, $\forall i \in \{1, \dots, k_n\}$, on a l'inégalité :

$$|\hat{E}_{n,\alpha}(u_i) - E_\alpha(u_i)| \leq \int_{-M}^{+M} |y|^\alpha dy \cdot \sup_{|y| \leq M} |\hat{h}_{n,u_i}(y) - h_{u_i}(y)|$$

d'où

$$\begin{aligned} \text{Prob} \left[\sup_{x \in [u, v]} |\hat{E}_{n,\alpha}(x) - E_\alpha(x)| > \varepsilon \right] &\leq \sum_{i=1}^{k_n} \text{Prob} \left[\sup_{|y| \leq M} |\hat{h}_{n,u_i}(y) - h_{u_i}(y)| \right. \\ &\quad \left. > \frac{\alpha+1}{4} \cdot \frac{\varepsilon}{M^{\alpha+1}} \right] \end{aligned}$$

en appliquant (26), on obtient, en posant $\frac{\alpha+1}{4} \cdot \frac{\varepsilon}{M^{\alpha+1}} = h$:

$$\text{Prob} \left[\sup_{x \in [u,v]} |\hat{E}_{n,\alpha}(x) - E_\alpha(x)| > \varepsilon \right] \leq \frac{v-u}{2a_n} K_1(h) \exp[-K_2(h) \frac{n a_n}{\alpha^2}]$$

$$\forall n > N(h)$$

et (36) permet de conclure.

2. Cas où Y n'est pas nécessairement borné.

Moyennant de légères modifications, la méthode précédente s'applique au cas où Y n'est pas nécessairement borné: on suppose que les hypothèses du théorème 20 et du corollaire 15 sont vérifiées et que :

$$E_\alpha(x) = \int_{\mathbf{R}} y^\alpha h_x(y) dy \quad \alpha > 0$$

existe, $\forall x \in [u,v]$. On définira alors un estimateur de $E_\alpha(x)$ en posant

$$E_{n,\alpha}^*(x) = \int_{-T_n}^{+T_n} y^\alpha \hat{h}_{n,u_i}(y) dy \quad \text{pour } x \in [u + (2i-2)a_n, u + 2ia_n]$$

$$i = 1, \dots, k_n - 1$$

et

$$E_{n,\alpha}^*(x) = \int_{-T_n}^{+T_n} y^\alpha \hat{h}_{n,u_{k_n}}(y) dy \quad \text{pour } x \in [u + 2(k_n-2)a_n, v]$$

on aura alors

THÉORÈME 21 bis :

Si

1°) Les hypothèses du théorème 20 et du corollaire 15 sont vérifiées.

$$2^\circ) \lim_{T \rightarrow \infty} \sup_{x \in [u,v]} \int_{|y| \geq T} |y|^\alpha h_x(y) dy = 0$$

$$3^\circ) T_n = n^\mu \text{ avec } 0 < \mu \frac{1}{\alpha+1} \cdot \inf(d, b(1-d), \frac{1-d}{2} - c)$$

alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in [u, v]} |E_{n, \alpha}^*(x) - E_{\alpha}(x)| = 0.$$

DÉMONSTRATION.

$$\text{Posons } E_{n, \alpha}(x) = \int_{-T_n}^{+T_n} y^{\alpha} h_x(y) dy.$$

D'après le 2°), il suffit de montrer que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in [u, v]} |E_{n, \alpha}^*(x) - E_{n, \alpha}(x)| = 0.$$

La démonstration est alors analogue à celle du théorème 21 : $E_{n, \alpha}(x)$ est Lipchitzienne d'ordre 1 et il existe une constante δ telle que :

$$|E_{n, \alpha}(x) - E_{n, \alpha}(u_i)| \leq \frac{2}{\alpha+1} \cdot \delta \cdot T_n^{\alpha+1} \cdot a_n$$

$$\forall x \in [u_i - a_n, u_i + a_n], \quad \forall i \in \{1, \dots, k_n\}.$$

$$\text{Comme on a } \mu < \frac{d}{\alpha+1}, \quad \text{il vient } \lim_{n \rightarrow \infty} T_n^{\alpha+1} \cdot a_n = 0$$

On a donc, pour n assez grand et ε positif donné, une inégalité analogue à (37), c'est-à-dire que :

$$\begin{aligned} \text{Prob} \left[\sup_{x \in [u, v]} |E_{n, \alpha}^*(x) - E_{n, \alpha}(x)| > \varepsilon \right] \\ \leq \text{Prob} \left[\sup_{1 \leq i \leq k_n} |E_{n, \alpha}^*(u_i) - E_{n, \alpha}(u_i)| > \frac{\varepsilon}{2} \right] \end{aligned}$$

Or, $\forall i \in \{1, \dots, k_n\}$ on a l'inégalité

$$|E_{n, \alpha}^*(u_i) - E_{n, \alpha}(u_i)| \leq \int_{-T_n}^{+T_n} |y|^{\alpha} dy \cdot \sup_{|y| \leq T_n} |\hat{h}_{n, u_i}(y) - h_{u_i}(y)|$$

d'où :

$$\begin{aligned} & \text{Prob} \left[\sup_{x \in [u, v]} |E_{n, \alpha}^*(x) - E_{n, \alpha}(x)| > \varepsilon \right] \\ & \leq \sum_{i=1}^k \text{Prob} \left[\sup_{y \in \mathbf{R}} |\hat{h}_{n, u_i}(y) - h_{u_i}(y)| > \frac{\varepsilon (\alpha+1)}{4 \cdot T_n \alpha+1} \right] \end{aligned} \quad (38)$$

Or, l'examen de la démonstration du corollaire 15 montre que pour $0 < \delta < \inf(d, b(1-d), \frac{1-d}{2} - c)$, on a une inégalité de la forme

$$\sup_{x \in [u, v]} \cdot \text{Prob} [n^\delta \cdot d(\hat{h}_{n, x}; h_x) > \varepsilon] \leq C_1 \exp [-C_2 n^\beta] \quad (39)$$

où C_1, C_2 et β sont positifs.

Alors (38) et (39) entraînent ,

$$\text{Prob} \left[\sup_{x \in [u, v]} |E_{n, \alpha}^*(x) - E_{n, \alpha}(x)| > \varepsilon \right] \leq \frac{v-u}{2a_n} C_1 \exp [-C_2 n^\beta] \quad (40)$$

car $0 < \mu (\alpha+1) < \inf(d, b(1-d), \frac{1-d}{2} - c)$.

Or, on a $a_n = C'_3 \cdot n^{-d}$ ($d > 0$), donc le deuxième membre de (40) est le terme général d'une série convergente et la démonstration est terminée.

REMARQUE.

Dans [36], où seul le cas $\alpha = 1$ est traité, les conditions imposées à la fonction de régression sont un peu moins fortes que les nôtres (continuité au lieu de conditions du type Lipchitz); en revanche, les conditions d'existence de moments conditionnels jusqu'à l'ordre 3 qui figurent dans [36] ne sont pas exigées ici.

3. Un cas particulier important : le régressogramme.

TUKEY dans [8] a proposé d'estimer la courbe de régression sur un intervalle compact $[u, v]$ à l'aide du *régressogramme*. Cet estimateur est

défini de la manière suivante: on fait une partition de $[u, v]$ en intervalles de même longueur et on estime la fonction de régression par la fonction en escalier qui, dans chaque intervalle de la partition, est constante et égale à la moyenne des ordonnées des points de l'échantillon dont l'abscisse appartient à l'intervalle considéré ⁽¹⁾. TUKEY pose alors le problème du choix de la partition: le théorème 21 permet de répondre partiellement à cette question; en effet, le régressogramme appartient à la famille d'estimateurs de la régression que nous venons d'étudier.

Pour le voir, il suffit d'estimer $h_x(y)$ à l'aide d'un noyau K possédant les propriétés suivantes: il existe un nombre positif A tel que $K(t) = 0$ pour $|t| \geq A$ et tel que:

$$\int_{-A}^{+A} K(t) dt = 1 \quad (41) \quad \int_{-A}^{+A} t K(t) dt = 0 \quad (42)$$

L'estimateur de $E_1(x)$ sera alors défini par

$$E_{n,1}^*(x) = 0, \quad \text{si } l_n = 0$$

$$= \frac{\alpha_{l_n}}{l_n} \sum_{k=1}^{l_n} \int_{-T_n}^{+T_n} y K[\alpha_{l_n}(y - y_{i_k})] dy, \quad \text{si } l_n \geq l.$$

D'où en faisant les changements de variables $t_{i_k} = \alpha_{l_n}(y - y_{i_k})$,
 $k = 1, \dots, l_n$.

$$E_{n,1}^*(x) = 0, \quad \text{si } l_n = 0$$

$$= \frac{1}{l_n} \sum_{k=1}^{l_n} y_{i_k} \int_{\alpha_{l_n}(-T_n - y_{i_k})}^{\alpha_{l_n}(T_n - y_{i_k})} K(t_{i_k}) dt_{i_k} -$$

$$- \frac{1}{l_n \cdot \alpha_{l_n}} \sum_{k=1}^{l_n} \int_{\alpha_{l_n}(-T_n - y_{i_k})}^{\alpha_{l_n}(T_n - y_{i_k})} t_{i_k} K(t_{i_k}) dt_{i_k}$$

si $l_n \geq l$.

(1) Si aucun point de l'échantillon n'est dans ce cas, on prend 0 comme valeur de l'estimateur sur l'intervalle considéré.

Comme $\lim_{n \rightarrow \infty} \alpha_n = +\infty$ on pourra toujours choisir $\alpha_n \geq 1$, $\forall n \in \mathbb{N}$; alors, si l'on suppose $|\mathcal{Y}| \leq M$ et si l'on tient compte du fait que $\lim_{n \rightarrow \infty} T_n = +\infty$, il vient, pour n assez grand :

$$\alpha \ell_n (T_n - y_{i_k}) \geq T_n - M > A$$

et

$$\alpha \ell_n (-T_n - y_{i_k}) \leq -T_n + M < A$$

d'où, en tenant compte de (41) et (42) :

$$\begin{aligned} E_{n,1}^*(x) &= 0 & \text{si } \ell_n &= 0 \\ &= \frac{1}{\ell_n} \sum_{k=1}^{\ell_n} y_{i_k} & \text{si } \ell_n &\geq 1 \end{aligned}$$

On obtient donc le régressogramme associé à la partition de $[u,v]$ en intervalles de longueur $2a_n$. Remarquons que l'estimateur obtenu ne dépend pas de la suite α_n . On aura donc le

COROLLAIRE 16 :

Sous les hypothèses du théorème 21 bis, le régressogramme associé aux intervalles de longueur $O(n^{-d})$, $0 < d < 1$, converge presque sûrement vers $E_1(x)$, uniformément pour $x \in [u,v]$.

V

ESTIMATION DE LA FONCTION FRONTIÈRE

I - POSITION DU PROBLÈME.

1. Introduction.

Dans ce chapitre, ainsi que dans le chapitre VI, nous étudierons des questions liées plus ou moins directement à l'estimation du *support* ⁽¹⁾ d'une loi de probabilité.

RENYI et SULANKE [38] ont étudié le cas où le support est un sous-ensemble convexe de $\mathbf{R}^2$; l'enveloppe convexe de l'échantillon est alors un estimateur naturel de la frontière du support.

Le problème, beaucoup plus compliqué dans le cas non-convexe, n'est pas résolu dans toute sa généralité. Cependant, M. GEFFROY [39] a traité un cas particulier important: nous donnons ci-dessous quelques extensions de certains de ses résultats, notamment celles que nous avons indiquées dans [40].

2. Fonction frontière.

Soit K un espace topologique compact, à base dénombrable, muni de sa tribu Borélienne $\mathcal{B}$ et soit λ une mesure de probabilité sur $(K, \mathcal{B})$ qui charge tout ouvert de K . Soit, d'autre part, s une application de K dans $\mathbf{R}$, continue et positive. Alors l'ensemble

$$S = \{(x, y) \in K \times \mathbf{R} \mid 0 \leq y \leq s(x)\}$$

est compact et, si K est connu, S est complètement déterminé par s ; nous dirons que s est la *fonction frontière* associée à S .

Soit, enfin, $\mathcal{D}$ l'ensemble des mesures de probabilité sur $(K \times \mathbf{R}, \mathcal{B} \otimes \mathcal{B}_{\mathbf{R}})$ qui ont pour support S et qui chargent tout ouvert de S . On se propose d'estimer s à partir d'un échantillon $(x_i, y_i)_{1 \leq i \leq n}$ tiré suivant $P \in \mathcal{D}$.

(1) Ou de sa frontière, ce qui est équivalent dans les cas usuels.

II - CONSTRUCTION D'UN ESTIMATEUR CONVERGENT.

1. Construction d'un estimateur.

Généralisant la méthode employée dans [39], nous allons construire une suite d'estimateurs de s à partir d'une suite de partitions de K .

LEMME 20 :

Si λ est non-atomique, il existe une suite (P_n) de partitions finies de K formées de n ouverts quarrables $Q_1^n, \dots, Q_n^n$ et d'un ensemble de mesure nulle N_n et telle que :

$$\lim_{n \rightarrow \infty} \max_{1 \leq \ell \leq n} |\lambda(Q_\ell^n)| = 0 \quad (1)$$

DÉMONSTRATION.

Tout d'abord, $\forall \varepsilon > 0$, il existe une partition finie de K formée d'ouverts quarrables de λ -mesure $\leq \varepsilon$ et d'un ensemble de λ -mesure nulle ⁽¹⁾.

Soit $n \in \mathbb{N}$; il existe donc une partition de K formée d'ouverts quarrables $Q_1^n, \dots, Q_{m_n}^n$ et d'un ensemble de mesure nulle N_n et telle que

$$\lambda(Q_i^n) \leq \frac{e^{-n}}{n}, \quad \forall i \in \{1, \dots, m_n\}$$

On peut toujours supposer que $\lambda(Q_i^n) > 0$, $\forall i \in \{1, \dots, m_n\}$.

Soit alors h_1 l'entier tel que

$$\lambda(Q_1^n + \dots + Q_{h_1-1}^n) < \frac{1}{n}$$

$$\text{et} \quad \lambda(Q_1^n + \dots + Q_{h_1}^n) \geq \frac{1}{n}$$

Posons $Q_1^n = Q_1^n + \dots + Q_{h_1}^n$; on aura $\frac{1}{n} \leq \lambda(Q_1^n) \leq \frac{1}{n} + \frac{e^{-n}}{n}$.

(1) cf. BOURBAKI [41], § 5.

Soit ensuite h_2 l'entier tel que

$$\lambda (Q'_{h_1+1} + \dots + Q'_{h_2-1}) < \frac{1}{n}$$

et

$$\lambda (Q'_{h_1+1} + \dots + Q'_{h_2}) \geq \frac{1}{n}$$

Posons de même $Q_2^n = Q'_{h_1+1} + \dots + Q'_{h_2}$; on aura

$$\frac{1}{n} \leq \lambda (Q_2^n) \leq \frac{1}{n} + \frac{e^{-n}}{n} .$$

En continuant le processus, on définira des ensembles Q_ℓ^n ($1 \leq \ell \leq n-1$) tels que $\frac{1}{n} \leq \lambda (Q_\ell^n) \leq \frac{1}{n} + \frac{e^{-n}}{n}$.

Enfin on peut poser

$$Q_n^n = \sum_{i=1}^{m_n} Q_i'^n = \sum_{\ell=1}^{n-1} Q_\ell^n$$

on aura alors

$$\frac{1}{n} - (n-1) \frac{e^{-n}}{n} \leq \lambda (Q_n^n) < \frac{1}{n} .$$

La partition $\{Q_1^n, \dots, Q_n^n; N_n\}$ répond à la question. En effet, les Q_ℓ^n sont bien des ouverts quarrables disjoints tels que

$$\sum_{\ell=1}^n \lambda (Q_\ell^n) = 1 \quad \text{et l'inégalité}$$

$$\max_{1 \leq \ell \leq n} \left| \frac{1}{n} - \lambda (Q_\ell^n) \right| \leq e^{-n}$$

entraîne bien (1) . C.Q.F.D.

DÉFINITION DE L'ESTIMATEUR.

Soit $k = k(n) \in \mathbf{N}^{(1)}$ et soit $\{N_1^k, \dots, N_k^k\}$ une partition de N_k en ensembles mesurables. Posons :

$$Q_{kj} = Q_j^k + N_j^k \quad j = 1, \dots, k$$

De même que M. GEFFROY dans [39] on définira l'estimateur de s par les relations :

$$\hat{s}_n(x) = \sup_{i \in J_{nj}} y_i, \quad \forall x \in Q_{kj}, \quad \forall j \in \{1, \dots, k\} \quad (3)$$

où J_{nj} est l'ensemble des indices $i \in \{1, \dots, n\}$ tels que $x_i \in Q_{kj}$ et où l'on pose conventionnellement

$$\sup_{i \in J_{nj}} y_i = 0 \quad \text{si } J_{nj} = \emptyset.$$

REMARQUE :

Comme, d'après le lemme 20, les ensembles Q_{kj}^0 forment, à un ensemble de mesure nulle près, une partition de K , on en déduit que $\hat{s}_n(x)$ est λ -presque partout continue.

NOTATIONS :

De même que dans le cas traité par M. GEFFROY, on peut poser :

$$M_{kj} = \max_{x \in \overline{Q_{kj}}} s(x) = \sup_{x \in Q_{kj}} s(x); \quad j = 1, \dots, k.$$

$$m_{kj} = \min_{x \in \overline{Q_{kj}}} s(x) = \inf_{x \in Q_{kj}} s(x), \quad j = 1, \dots, k.$$

$$S_n = \max_{1 \leq j \leq k} (M_{kj} - U_{nj})$$

(1) On suppose que $\lim_{n \rightarrow \infty} k(n) = +\infty$.

$$\text{avec } U_{nj} = \sup_{i \in J_{nj}} y_i$$

et on peut considérer les ensembles :

$$D_{kj} = \{ (x, y) \in K \times \mathbf{R} \mid x \in Q_{kj}, \quad 0 \leq y \leq s(x) \}$$

$$D_{kj}^c(z) = \{ (x, y) \in K \times \mathbf{R} \mid x \in Q_{kj}, \quad s(x) \geq M_{kj} - z \}$$

$$j = 1, \dots, k.$$

Enfin, si l'on désigne par $E_{nj}(z)$ l'évènement

$$\bigcap_{i=1}^n \{ (x_i, y_i) \in D_{kj}^c(z) \}$$

et par $E'_{nj}(z)$ l'évènement contraire on aura encore

$$\text{Prob} [Z_n < z] = \text{Prob} \left[\bigcap_{j=1}^k E'_{nj}(z) \right].$$

2. Convergence de l'estimateur.

Le théorème 2 de [39] se généralise de la façon suivante :

THÉORÈME 22 :

Si $P \in \mathcal{D}$ est absolument continue par rapport à $\lambda \otimes \ell$ (ℓ désignant la mesure de Lebesgue sur $\mathbf{R}$), et s'il existe une version de sa dérivée de Radon-Nikodym, soit $\varphi(x, y)$ telle que l'on ait :

$$A \leq \varphi(x, y) \leq B, \quad \forall (x, y) \in S \quad (4)$$

où A et B sont deux constantes positives, alors, pour que $\hat{s}_n$ tende vers s uniformément presque sûrement, il est nécessaire que $k(n) = o\left(\frac{n}{\text{Log } n}\right)$.

Si de plus, le maximum des diamètres des ensembles Q_{kj} (relatif à une distance compatible avec la topologie de K) tend vers 0, la condition est aussi suffisante.

DÉMONSTRATION.

Soit d la distance de la convergence uniforme.

On a successivement

$$d(\hat{s}_n, s) \xrightarrow[p.s.]{} 0 \implies Z_n \xrightarrow[p.s.]{} 0 \implies \text{Prob} \left[\bigcup_{j=1}^k E_{nj}(z) \right] \rightarrow 0.$$

D'autre part

$$\begin{aligned} \text{Prob} [D_{kj}(z)] &= \iint_{D_{kj}(z)} \varphi(x, y) d\lambda(x) d\ell(y) \\ &\leq \iint_{Q_{kj} \times [M_{kj}^{-s}, M_{kj}]} \varphi(x, y) d\lambda(x) d\ell(y) \\ &\leq B \int_{Q_{kj}} d\lambda(x) \int_{M_{kj}^{-z}}^{M_{kj}} d\ell(y) \end{aligned}$$

$$\text{d'où } \text{Prob} [D_{kj}(z)] \leq Bz \lambda(Q_{kj}) \leq \frac{Bz}{k} (1 + k e^{-k})$$

la dernière inégalité étant une conséquence de (2).

On aura donc, pour k assez grand

$$\text{Prob} [D_{kj}(z)] < \frac{Bz}{k} (1 + \varepsilon) \quad (\varepsilon > 0, \text{ donné})$$

et l'on termine comme dans [39] (p. 197-98).

RÉCIPROQUE :

$$\text{Soit } z \leq \frac{1}{2} \left[\inf_{x \in K} s(x) \right] = \frac{m}{2}. \text{ On aura alors}$$

$$\text{Prob} [D_{kj}(z)] = \iint_{D_{kj}(z)} \varphi(x, y) d\lambda(x) d\ell(y) \geq A \frac{m}{2} \lambda(Q_{kj})$$

donc, d'après (2), on aura

$$\text{Prob} [D_{kj}(z)] \geq \frac{Am}{2k} (1 - k e^{-k}) \geq \frac{Am}{2k} (1 - \varepsilon) \quad \text{pour } k \text{ assez grand.}$$

Alors, de même que dans [39] (p. 198-99), on en déduit que $Z_n \xrightarrow[p]{p} 0$ puis que $Z_n \xrightarrow[p.s.]{p} 0^{(1)}$.

Enfin, compte tenu de l'inégalité

$$d(\hat{s}_n, s) \leq Z_n + \sup_{1 \leq j \leq k} (M_{kj} - m_{kj})$$

et de l'hypothèse faite sur les diamètres des Q_{kj} , on en déduit que

$$\lim_{n \rightarrow \infty} p.s. d(\hat{s}_n, s) = 0.$$

C.Q.F.D.

REMARQUE:

La convergence est même presque complètement sûre.

III - CAS D'UN ESPACE LOCALEMENT COMPACT.

Les résultats précédents s'étendent facilement au cas où K est localement compact à base dénombrable ⁽²⁾. On reprend les mêmes hypothèses que dans 1.2 et on suppose, en outre, que s admet, suivant le filtre des voisinages de l'infini, une limite σ positive.

Soit alors K' le compactifié d'Alexandroff de K : $K' = K + \{\omega\}$, ω désignant le point à l'infini de K . On peut se ramener au problème précédent en posant :

$$\begin{cases} s'(x) = s(x) & \text{pour } x \in K \\ s'(\omega) = \sigma \end{cases}$$

$$\begin{cases} \lambda'(A) = \lambda(A) & \text{si } A \text{ est un Borélien de } K \\ \lambda'(\{\omega\}) = 0 \end{cases}$$

et, $\forall P \in \mathcal{D}$,

(1) D'après la démonstration du corollaire 2 de [39] (p.199).

(2) K est supposé non compact.

$$\begin{cases} P'(A) = P(A) & \text{si } A \text{ est un Borélien de } K \\ P'(\{\omega\}) = 0 \end{cases}$$

Le lemme 20 s'applique alors à λ' donc, par projection sur K , à λ et, $\hat{s}'_n$ désignant l'estimateur de s' défini par (3), on voit donc que le théorème 22 s'applique à $\hat{s}'_n$ donc à sa restriction à K , $\hat{s}_n$. Finalement, on peut énoncer le

COROLLAIRE 17 :

Si K est localement compact à base dénombrable, le lemme 20 et le théorème 22 restent valables.

REMARQUE.

Dans l'énoncé du théorème 22 appliqué au cas localement compact, la distance considérée doit être compatible avec la topologie de K' .

APPLICATION :

Le corollaire 17 s'applique au cas particulier important où $K = \mathbf{R}$.

Commençons par exhiber une famille de suites de partitions de $\mathbf{R}$ qui satisfont aux conditions du lemme 20 et du théorème 22. Nous nous appuierons sur le lemme suivant :

LEMME 21 :

Soit E un espace métrique compact et soit μ une mesure sur E qui charge tout ouvert, alors si à tout $n \in \mathbf{N}$ on associe des Boules $B_1, \dots, B_{h(n)}$ de telle façon que

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq h(n)} \mu(B_i) = 0 \quad (5)$$

on a aussi

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq h(n)} D(B_i) = 0 \quad (6)$$

$D(B_i)$ désignant le diamètre de B_i ; $i = 1, \dots, h(n)$.

DÉMONSTRATION.

On notera $B(a, r)$ la boule ouverte de centre a et de rayon r . Soit $B(y_n, r_n)$ une suite de Boules telle que $\mu[B(y_n, r_n)] \rightarrow 0$. Si $r_n \not\rightarrow 0$, il existe $\varepsilon > 0$ et une suite extraite (r_{n_k}) telle que $r_{n_k} \geq \varepsilon$, $\forall k \in \mathbb{N}$.

On a alors

$$\mu[B(y_{n_k}, r_{n_k})] \geq \mu[B(y_{n_k}, \varepsilon)] \quad , \quad \forall k \in \mathbb{N}.$$

donc

$$\lim_{n \rightarrow \infty} \mu[B(y_{n_k}, \varepsilon)] = 0.$$

Or, de (y_{n_k}) on peut extraire une suite $(y_{n_{k_\ell}})$ qui converge vers un élément de y de E donc, pour ℓ assez grand on aura

$$B(y, \frac{\varepsilon}{2}) \subset B(y_{n_{k_\ell}}, \varepsilon)$$

mais d'après l'hypothèse faite sur μ , $\mu[B(y, \frac{\varepsilon}{2})] > 0$ et l'on aboutit à une contradiction. On a donc nécessairement $r_n \rightarrow 0$.

Cela dit, la démonstration du lemme 21 s'achève de la façon suivante: Parmi les Boules $B_1, \dots, B_{h(n)}$, soit B_n celle qui a le plus petit indice parmi celles de plus grand diamètre; d'après (5), on a $\mu(B_n) \rightarrow 0$ donc, d'après ce qui précède, le rayon de B_n tend vers 0, ce qui entraîne immédiatement (6). C.Q.F.D.

Soit alors λ une mesure de probabilité sur $\mathbf{R}^{(1)}$ dont la fonction

(1) On posera ici $\bar{\mathbf{R}} = \mathbf{R} + \{-\infty, +\infty\}$: il est clair que le corollaire 17 s'applique encore moyennant les hypothèses

$$\lim_{x \rightarrow -\infty} s(x) = \sigma > 0 \quad ; \quad \lim_{x \rightarrow +\infty} s(x) = \sigma' > 0.$$

de répartition F_λ est continue et strictement croissante; on peut lui associer la suite de partitions :

$$[-\infty, a_1[; [a_1, a_2[; \dots [a_{k-2}, a_{k-1}[; [a_{k-1}, +\infty] \quad (7)$$

$$\text{avec } a_i = F_\lambda^{-1}\left(\frac{i}{k}\right); \quad i = 1, \dots, k-1; \quad k = 1, 2, \dots \quad (8)$$

Il est clair que la condition (5) est vérifiée donc, d'après le lemme 21, la condition (6) est également vérifiée.

REMARQUE :

Dans (6), D désigne le diamètre relatif à une distance d compatible avec la topologie de $\mathbf{R}$ et pour laquelle les ensembles de la partition précédente ont pour intérieur des Boules ouvertes. Il est facile de construire une telle distance. On pourra prendre, par exemple

$$d(x, y) = |\text{Arc tg } x - \text{Arc tg } y| \quad ; \quad x, y \in \mathbf{R}.$$

CONCLUSION.

Si P est une mesure de probabilité sur $\mathbf{R}^2$ dont le support est limité par la droite $y = 0$ et le graphe d'une fonction $s(x)$ continue, strictement positive et admettant des limites positives pour $x \rightarrow +\infty$; si, de plus, P est absolument continue par rapport à $\lambda \otimes \ell$ et si sa densité vérifie (4), alors le corollaire 17 s'applique, c'est-à-dire que, pour

$$k(n) = O\left(\frac{n}{\text{Log } n}\right) \text{ on a :}$$

$$\lim_{n \rightarrow \infty} \text{p.s. } d(\hat{s}_n, s) = 0$$

$\hat{s}_n$ étant l'estimateur défini par (3) associé à la partition définie par (7) et (8).

IV - CAS OU LA FONCTION FRONTIÈRE EST DE SIGNE QUELCONQUE.

1. Préliminaires.

Dans ce paragraphe, nous nous proposons de montrer que les résul-

tats précédents subsistent quand $s(x)$ peut s'annuler ou changer de signe un nombre fini de fois sur $\mathbf{R}$.

LEMME 22 :

Soit (X_n) une suite de variables aléatoires telle que :

$$\lim_{n \rightarrow \infty} p. X_n = 0$$

Soit (N_k) une suite de variables aléatoires à valeurs entières telle que :

$$\lim_{n \rightarrow \infty} p. N_k = \infty$$

Alors, si les suites (X_n) et (N_k) sont indépendantes, on a aussi

$$\lim_{n \rightarrow \infty} p. X_{n_k} = 0$$

DÉMONSTRATION.

Soit $x > 0$; on peut écrire $\forall k \in \mathbf{N}$

$$\text{Prob} [|X_{N_k}| > x] = \sum_{n \in A_k} \text{Prob} [N_k = n] \text{Prob} [|X_{N_k}| > x | N_k = n]$$

avec $A_k = \{n \in \mathbf{N} | \text{Prob} [N_k = n] > 0\}$

Compte tenu de l'hypothèse d'indépendance, il vient :

$$\begin{aligned} \text{Prob} [|X_{N_k}| > x] &= \sum_{n \in A_k} \text{Prob} [N_k = n] \cdot \text{Prob} [|X_n| > x] \\ &= \sum_{n=0}^{+\infty} \text{Prob} [N_k = n] \cdot \text{Prob} [|X_n| > x] \end{aligned}$$

Alors, $\forall \varepsilon > 0$, $N(\varepsilon, x)$ tel que $n \geq N(\varepsilon, x) \Rightarrow \text{Prob} [|X_n| > x] < \frac{\varepsilon}{2}$

d'où :

$$\text{Prob} [|X_{N_k}| > x] < \sum_{n=0}^{N(\varepsilon, x)} \text{Prob} [N_k = n] + \frac{\varepsilon}{2}$$

mais $\exists [K \quad [N(\varepsilon, x)] \quad \text{tel que}$

$$k \geq K [N(\varepsilon, x)] \implies \text{Prob} [N_k \leq N(\varepsilon, x)] < \frac{\varepsilon}{2}$$

d'où finalement

$$\text{Prob} [|X_{N_k}| > x] < \varepsilon \quad \text{dès que } k \geq K [N(\varepsilon, x)]$$

$\forall x > 0, \quad \forall \varepsilon > 0. \quad \text{C.Q.F.D.}$

NOTATIONS

$\hat{s}_n$ désignant l'estimateur défini par (3), nous dirons dorénavant que $\hat{s}_n$ est l'estimateur de Geffroy associé à la partition $\{Q_{kj}\}$.

D'autre part, nous supposons que K est un intervalle de $\mathbf{R}$. Soit alors J un sous-intervalle de K qui est réunion d'ensembles Q_{kj} .

$$\text{Posons } S_J = \{(x, y) \in J \times \mathbf{R} \mid 0 \leq y \leq s(x)\}$$

et désignons par s_J et $\hat{s}_{n,J}$ les restrictions respectives de s et de $\hat{s}_n$ à J .

Enfin P étant une mesure de probabilité sur $\mathbf{R}^2$, appelons P_J la mesure de probabilité induite par P sur S_J .

On a alors le

LEMME 23 :

Si les points de l'échantillon sont tirés suivant la loi de probabilité P , $\hat{s}_{n,J}$ est l'estimateur de Geffroy de s_J associé à la partition induite par $\{Q_{kj}\}$ sur J , à la loi de probabilité P_J et à un échantillon dont la taille est une variable aléatoire indépendante des variables de l'échantillon et qui suit une loi binominale de paramètres $P(S_J)$ et n .

DÉMONSTRATION.

Posons $J = \sum_{j=1}^{k_1} Q_{kj}$ (ce que l'on peut faire au besoin en changeant les notations) et

$$D'_{kj}(z) = \{ (x,y) \in Q_{kj} \times \mathbf{R} \mid z < y < f(x) \} ,$$

$$\forall z \in [0, f(x)] , \quad \forall x \in J , \quad \forall j \in \{1, \dots, k_1\} .$$

On aura successivement et $\forall j \in \{1, \dots, k_1\}$:

$$\begin{aligned} \text{Prob} [U_{nj} < z] &= \text{Prob} [\bigcap_{i=1}^n (x_i, y_i) \notin D'_{kj}(z)] \\ &= [1 - P(D'_{kj}(z))]^n \\ &= [1 - P(S_j) + P(S_j) - P(D'_{kj}(z))]^n \\ &= \sum_{\ell=0}^n C_n^\ell [1 - P(S_j)]^{n-\ell} \left[\frac{P(S_j - D'_{kj}(z))}{P(S_j)} \right]^\ell \\ &= \sum_{\ell=0}^n \text{Prob} [L_n = \ell] P_J^{\otimes \ell} \left[\bigcap_{i=1}^\ell (x_i, y_i) \notin D'_{kj}(z) \right] \end{aligned}$$

où L_n désigne une variable aléatoire qui suit une loi binominale de paramètres $P(S_j)$ et n .

Cette dernière expression de $\text{Prob} [U_{nj} < z]$ correspond bien aux conditions de l'énoncé du lemme 23 .

2. Cas où s s'annule à l'infini.

COROLLAIRE 18 :

Si s est continue, positive, définie $\forall x \in \mathbf{R}$ et telle que $\lim_{|x| \rightarrow \infty} s(x) = 0$, si, d'autre part, P est absolument continue par rapport à la mesure de Lebesgue sur $\mathbf{R}^2$ et s'il existe une version de sa densité,

soit $\varphi(x, y)$, telle que, $\forall I$ intervalle borné de $\mathbf{R}$ il existe deux constantes positives $A(I)$ et $B(I)$ telles que

$$A(I) \leq \varphi(x, y) \leq B(I) \quad , \quad \forall (x, y) \in S_I \quad (9)$$

alors, pour que l'estimateur de Geffroy associé aux ensembles

$$\left[p + \frac{j-1}{k}, p + \frac{j}{k} \right[\quad , \quad p \in \mathbf{Z} \quad ,$$

$j = 1, \dots, k$, converge uniformément en probabilité, il suffit que

$$k(n) = O\left(\frac{n}{\text{Log } n}\right)^{(1)} .$$

DÉMONSTRATION.

Soit $\varepsilon > 0$, il existe un entier A tel que $|x| \geq A$ entraîne $s(x) < \varepsilon$.

Soient s_A et $\hat{s}_{n,A}$ les restrictions de s et de $\hat{s}_n$ à $[-A, +A]$: d'après le lemme 23, $\hat{s}_{n,A}$ est l'estimateur de Geffroy associé à la partition

$$\left[p + \frac{j-1}{k}, p + \frac{j}{k} \right[\quad , \quad p \in \{-A, -A+1, \dots, A-1\} \quad j \in \{1, \dots, k\} ,$$

à la loi $\frac{P}{P(S_{[-A, +A]})}$, et à un échantillon dont la taille L_n est une variable aléatoire qui suit une loi binominale de paramètres $P(S_{[-A, +A]})$ et n .

On a donc $\lim_{n \rightarrow \infty} p \cdot L_n = +\infty$.

Or, d'après le théorème 22 :

$$k = O\left(\frac{n}{\text{Log } n}\right) \implies 2A k = O\left(\frac{n}{\text{Log } n}\right) \implies \lim_{q \rightarrow \infty} p \cdot d(s_A; \hat{s}'_{q,A}) = 0$$

(1) Remarquons que, dans ce corollaire, λ n'est plus une mesure bornée mais est la mesure de Lebesgue sur $\mathbf{R}$ et que $\{Q_{kj}\}$ est formée d'une infinité d'intervalles.

$\hat{s}'_{q,A}$ désignant ici l'estimateur de Geffroy de s_A associé à la partition précédente et à un échantillon de taille q tiré suivant $P_{[-A, +A]}$.

D'où, en appliquant le lemme 22 aux suites

$d(s_A; \hat{s}'_{q,A})$ et L_n :

$$k = O\left(\frac{n}{\log n}\right) \implies \lim_{n \rightarrow \infty} p. d(s_A; \hat{s}'_{n,A}) = 0$$

Enfin, l'équivalence $d(s_A; \hat{s}'_{n,A}) < \varepsilon \iff d(s, \hat{s}_n) < \varepsilon$

$\forall n \in \mathbb{N}$ permet de conclure.

3. Cas où s s'annule à distance finie.

Dans cette partie, nous supposons que K est un intervalle compact de $\mathbb{R}$; on pourra prendre $K = [0, 1]$ sans restreindre la généralité.

LEMME 24 :

Soit $\{Q_{kj}^{r_0}\}_{j=1, \dots, h}$ la partition induite par

$\left[\frac{j-1}{k}, \frac{j}{k}\right]$, $j = 1, \dots, k-1$ sur l'intervalle $[0, r_0]$ où r_0 est un rationnel appartenant à $]0, 1[$. Alors les longueurs des intervalles $Q_{kj}^{r_0}$, sont des infiniment petits de même ordre que $\frac{1}{k}$.

DÉMONSTRATION.

Posons $r_0 = \frac{p_0}{q_0}$ en supposant cette fraction irréductible. h désignant le nombre d'intervalles de la partition $\{Q_{kj}^{r_0}\}$, il est clair que $r_0 k \leq h \leq r_0 k + 1$; h est donc un infiniment grand du même ordre que k .

Si $r_0 k$ est entier, donc si $k = \lambda q_0$, $\lambda \in \mathbb{N}$, on a $h = r_0 k$ et la partition est formée de h intervalles de longueur $\frac{1}{k}$.

Si $r_0 k$ n'est pas entier, on peut écrire :

$$k = \lambda q_0 + \mu, \quad \lambda \in \mathbb{N}, \quad \mu \in \mathbb{N}, \quad \text{avec } 1 \leq \mu \leq q_0 - 1$$

et l'on a alors $h = [r_0 k + 1] = [r_0 k] + 1$: dans ce cas, les $h-1$ premiers intervalles de la partition $\{Q_{kj}^{r_0}\}$ sont de longueur $\frac{1}{k}$; montrons que $[\frac{h-1}{k}, r_0]$ a une longueur de l'ordre de $\frac{1}{k}$. On peut écrire

$$r_0 - \frac{h-1}{k} = \frac{1}{k} [r_0 k - [r_0 k]]$$

$$\text{Or } r_0 k - [r_0 k] = \frac{P_0}{q_0} (\lambda q_0 + \mu) - [\frac{P_0}{q_0} (\lambda q_0 + \mu)]$$

$$= r_0 \mu - [r_0 \mu]$$

Or $r_0 \mu$ ne peut être entier puisque $1 \leq \mu \leq q_0 - 1$, donc $r_0 \mu - [r_0 \mu] \neq 0$ et comme cette quantité ne peut prendre qu'un nombre fini de valeurs, la démonstration est terminée.

COROLLAIRE 19 :

Si s est continue sur $[0,1]$ et positive sauf au point $c \in]0,1[$ où elle s'annule, l'estimateur de Geffroy associé à la partition

$$G_k = \{ [0, \frac{1}{k}[, [\frac{1}{k}, \frac{2}{k}[, \dots, [\frac{k-1}{k}, 1] \}$$

converge uniformément en probabilité si $k(n) = O(\frac{n}{\text{Log } n})$ et si la condition (9) est vérifiée pour $I = [0,1]$.

DÉMONSTRATION.

Soit $\varepsilon > 0$; on peut trouver deux rationnels r_0 et $r_1 \in]0,1[$ tels que $c \in]r_0, r_1[$ et $0 \leq s(x) \leq \frac{\varepsilon}{4}$, $\forall x \in [r_0, r_1]$.

Soit $G_k^{r_0}$ la partition induite par G_k sur $[0, r_0]$ et soit $\hat{t}_{n, r_0}$

l'estimateur de Geffroy associé à l'échantillon induit sur $S_{[0, r_0]}$. Un raisonnement tout à fait analogue à celui qui a été fait dans la démonstration du corollaire 18 montre, compte tenu du lemme 24, que $\hat{t}_{n, r_0}$ converge uniformément en probabilité vers $s_{[0, r_0]}$.

Or, $\hat{t}_{n, r_0} = \hat{s}_{n, [0, r_0]}$ sauf peut-être sur le $h^{\text{ème}}$ intervalle de G_k^{e} où l'on a $\hat{t}_{n, r_0} \leq \hat{s}_{n, [0, r_0]}$. Mais on peut choisir n donc k et donc h assez grands pour que, sur cet intervalle, on ait $s(x) < \frac{\varepsilon}{2}$, ce qui entraîne alors $d(\hat{t}_{n, r_0}; \hat{s}_{n, [0, r_0]}) < \frac{\varepsilon}{2}$.

On en déduit donc que $\hat{s}_{n, [0, r_0]}$ converge uniformément en probabilité vers $s_{[0, r_0]}$.

On montre de même que $\hat{s}_{n, [r_1, 1]}$ converge uniformément en probabilité vers $s_{[r_1, 1]}$.

Enfin, il est clair que $d(s_{[r_0, r_1]}, \hat{s}_{n, [r_0, r_1]}) < \frac{\varepsilon}{4}$.

Or, $d(\hat{s}_n, s) = \sup [d(s_{[0, r_0]}, \hat{s}_{n, [0, r_0]}); d(s_{[r_0, r_1]}, \hat{s}_{n, [r_0, r_1]});$

$$d(s_{[r_1, 1]}, \hat{s}_{n, [r_1, 1]})]$$

donc, pour n assez grand

$$\text{Prob} [d(\hat{s}_n, s) > \varepsilon] < \varepsilon$$

C.Q.F.D.

4. Cas où s change de signe.

Si s n'est plus supposée positive, il faut généraliser l'estimateur défini par (3). On posera :

$$\hat{v}_n(x) = \sup_{i \in J_{nj}^+} y_i + \inf_{i \in J_{nj}^-} y_i, \quad \forall x \in Q_{kj}, \quad \forall j$$

avec $J_{nj}^+ = \{i \in \{1, \dots, n\} \mid x_i \in Q_{kj}, y_i \geq 0\}$

$J_{nj}^- = \{i \in \{1, \dots, n\} \mid x_i \in Q_{kj}, y_i < 0\}$

et les conventions $\sup_{i \in J_{nj}^+} y_i = 0$ pour $J_{nj}^+ = \emptyset$

et $\inf_{i \in J_{nj}^-} y_i = 0$ pour $J_{nj}^- = \emptyset$.

On a alors le

COROLLAIRE 20 :

Sous les conditions du corollaire 18, excepté le fait que s peut s'annuler ou changer de signe un nombre fini de fois sur $\mathbf{R}$, $\hat{v}_n$ converge uniformément en probabilité vers s .

DÉMONSTRATION.

Immédiate en appliquant les corollaires 18 et 19 à $|s|$.

5. Convergence presque sûre.

Nous avons énoncé les résultats de ce paragraphe "en convergence en probabilité", mais ils restent valables "en convergence presque sûre"; c'est une conséquence du

LEMME 22 bis :

Soit (X_n) une suite de variables aléatoires telle que

$$\lim_{n \rightarrow \infty} \text{p.s. } X_n = 0$$

Soit (N_k) une suite de variables aléatoires à valeurs entières telle que

$$\lim_{k \rightarrow \infty} \text{p.s. } N_k = \infty$$

Alors, on a aussi

$$\lim_{k \rightarrow \infty} \text{p.s. } X_{N_k} = 0$$

DÉMONSTRATION.

D'après les hypothèses, il existe dans $\Omega^{\mathbb{N}}$ un ensemble E de probabilité nulle et tel que $X_n(\omega) \rightarrow 0$ et $N_k(\omega) \rightarrow +\infty$, $\forall \omega \in \Omega^{\mathbb{N}} - E$. Il s'agit de montrer que l'on a aussi $X_{N_k}(\omega) \rightarrow 0$, $\forall \omega \in \Omega^{\mathbb{N}} - E$.

Or, $\forall \omega \in \Omega^{\mathbb{N}} - E$, $\forall \varepsilon > 0$, il existe un entier $\lambda(\varepsilon, \omega)$ tel que $|X_n(\omega)| < \varepsilon$; il existe alors un entier $K(\varepsilon, \omega)$ tel que $k > K(\varepsilon, \omega)$ entraîne $N_k(\omega) > \lambda(\varepsilon, \omega)$, ce qui entraîne finalement $|X_{N_k}(\omega)| < \varepsilon$.
C.Q.F.D.

V - APPLICATION : DÉFINITION D'UNE MÉTHODE DE MONTE CARLO

Revenons au cas où s est une fonction continue et positive définie sur $[0,1]$. Sous les hypothèses du théorème 2 de [39], il est clair que :

$$\int_0^1 \hat{s}_n(x) dx \xrightarrow{\text{p.s.}} \int_0^1 s(x) dx \quad (10)$$

On peut utiliser (10) pour définir une Méthode de Monte Carlo :

Soit à calculer $\int_0^1 s(x) dx$ où s est une fonction continue et positive connue: on tire $2n$ nombres $x_1, \dots, x_n; z_1, \dots, z_n$ suivant une loi uniforme sur $[0,1]$ et indépendamment les uns des autres et on estime

$$\int_0^1 s(x) .dx$$

par

$$\int_0^1 \hat{s}_n(x) dx$$

$\hat{s}_n$ étant construit à partir de l'échantillon

$$(x_i, z_i s(x_i))_{1 \leq i \leq n}.$$

La précision de cette Méthode dépend évidemment du choix de $k(n)$ et on peut montrer le résultat suivant :

Si s est Lipchitzienne et si $k(n) = O(n^\alpha)$, $\frac{1}{3} < \alpha < \frac{1}{2}$,

$$\text{alors } \sigma^2 \left[\int_0^1 \hat{s}_n(x) dx \right] = O\left(\frac{1}{n^{1-\alpha}}\right)$$

VI

ÉLÉMENTS EXTRÊMAUX D'UN ÉCHANTILLON VECTORIEL

I - GÉNÉRALITES

1. Introduction.

Le problème que nous étudions dans ce chapitre nous a été proposé par M. GEFFROY: il s'agit d'étudier les éléments extrémaux d'un échantillon de variables aléatoires à valeurs dans un espace partiellement ordonné.

Cette étude semble intéressante car ces éléments sont analogues aux valeurs extrêmes d'un échantillon de variables aléatoires réelles. De plus, la répartition des éléments extrémaux donne des indications sur l'aspect du support de la loi des variables de l'échantillon.

Dans la suite, nous étudions surtout le comportement asymptotique du nombre d'éléments extrémaux: nous donnons les résultats qui figurent dans [42] ainsi que certains résultats inédits que M. GEFFROY a eu l'obligeance de nous communiquer.

2. Éléments extrémaux d'une partie d'un espace ordonné.

a) Soit E un espace ordonné. Nous noterons $\leq$ la relation d'ordre de E , $\geq$ la relation d'ordre inverse; enfin si x et y ne sont pas comparables, nous écrirons $x \wedge y$.

DÉFINITION 12 :

Soit A une partie de E . On dit que $y \in A$ est un élément **maximal** de A si, $\forall x \in A$, on a $x \leq y$ ou $x \wedge y$. On dit que $y \in A$ est un élément **minimal** de A si, $\forall x \in A$, on a $x \geq y$ ou $x \wedge y$. On dit que $y \in A$ est un élément **extrémal** de A si y est un élément maximal ou minimal de A .

L'existence d'éléments extrémaux est assurée par le

LEMME 25 :

Si A est une partie finie de E , alors A admet toujours au moins un élément maximal (resp. minimal).

DÉMONSTRATION.

Immédiate par récurrence sur le nombre d'éléments de A .

b)

LEMME 26 :

Soit f une application de E dans E . La relation R_f définie par $x R_f y \iff f(x) \leq f(y)$ est une relation d'ordre si et seulement si f est injective.

DÉMONSTRATION.

Elémentaire.

NOTATION :

$x R_f y$ sera notée $x \leq y (f)$ et appelée relation d'ordre induite par f sur E .

DÉFINITION 12 bis :

Soit A une partie de E . On dira que $y \in A$ est un élément extrémal de A , modulo f , si y est un élément extrémal de A pour la relation d'ordre R_f .

Propriétés élémentaires.

On montre très facilement les propriétés suivantes :

1°) R_f coïncide avec $\leq$ si et seulement si f est un automorphisme de E .

2°) Soient f et g deux bijections de E . f et g induisent le même ordre sur E si et seulement si $g \circ f^{-1}$ est un automorphisme de E .

3°) Si E est un treillis et si f est une bijection de E , alors E est un treillis modulo f et l'on a les relations :

$$\sup_f \{x, y\} = f^{-1} [\sup \{f(x), f(y)\}] \quad (1)$$

$$\inf_f \{x, y\} = f^{-1} [\inf \{f(x), f(y)\}] \quad (2)$$

4°) Si E est un groupe ordonné les applications f et $-f$ induisent des relations d'ordre inverses l'une de l'autre.

3. Éléments extrémaux d'un échantillon de variables aléatoires à valeurs dans un espace mesurable ordonné.

a) DÉFINITION 13 :

On dira que E est un espace mesurable ordonné (E.M.O.) si E est un espace mesurable, un espace ordonné et si les intervalles de E sont mesurables.

Si E est un E.M.O., si f est une application injective et mesurable de E dans E et si P est une mesure de probabilité sur E l'application F_f^P définie par

$$F_f^P(x) = P(\] \leftarrow, x[f) , \quad \forall x \in E \quad (3)$$

est la fonction de répartition de P , modulo f .⁽¹⁾

La définition de R_f montre que

$$F_f^P(x) = F_u^{fP}[f(x)] , \quad \forall x \in E \quad (4)$$

fP désignant la mesure de probabilité image de P par f et u désignant l'application identique de E .

b) Soit $x_1, \dots, x_n$ un échantillon de taille n à valeurs dans E , on dira que x_i est un élément maximal de l'échantillon si x_i est un élément maximal de l'ensemble $\{x_1, \dots, x_n\}$. On définit de même les éléments minimaux, extrémaux, extrémaux modulo f etc... de l'échantillon.

Remarquons qu'on ne peut pas définir les éléments maximaux comme des variables aléatoires à valeurs dans E , leur nombre n'étant pas fixe.

(1) $\] \leftarrow, x[f$ désigne l'ensemble des $y \in E$ tels que $y \leq x(f)$ et $y \neq x$.

Par contre, le nombre d'éléments maximaux de l'échantillon est une variable aléatoire réelle parfaitement définie.

II - LOI DU NOMBRE D'ÉLÉMENTS MAXIMAUX ASSOCIÉS A UNE ROTATION DE $\mathbf{R}^2$.

Nous commençons maintenant l'étude des propriétés du nombre d'éléments maximaux d'un échantillon. Nous étudierons essentiellement le cas où $E = \mathbf{R}^2$ muni de sa tribu Borélienne et de sa relation d'ordre habituelle.

NOTATIONS :

- $X_1, \dots, X_n$ désignent n vecteurs aléatoires à valeurs dans $\mathbf{R}^2$, indépendants et équidistribués, de loi commune P .

- θ désigne la rotation de centre 0 et d'angle θ .

- $Y_n(\theta)$ désigne le nombre d'éléments maximaux modulo θ de l'échantillon $X_1, \dots, X_n$.

- On pose :

$$\varphi_{\theta}^{P,k}(x_1, \dots, x_k) = P \left[\bigcup_{i=1}^k \left\{ x_i \in \theta \right\} \right] \quad (5)$$

$$\forall x_1, \dots, x_k \in \mathbf{R}^2 ; \quad \forall k \in \mathbf{N},$$

et

$$A_k(\theta) = \{ (x_1, \dots, x_k) \in \mathbf{R}^{2k} \mid (i_1, \dots, i_k) \text{ permutation de } (1, \dots, k) \text{ telle que } x_{i_1} \leq \dots \leq x_{i_k} \left(\theta + \frac{\pi}{2} \right) \} \quad (6)$$

LEMME 27 :

On a les relations suivantes :

$$1^0) \quad \varphi_{\theta}^{P,k}(x_1, \dots, x_k) = \varphi_0^{\theta P,k}(\theta x_1, \dots, \theta x_k) \quad (7)$$

$$\forall x_1, \dots, x_k \in \mathbf{R}^2, \quad \forall k \in \mathbf{N}.$$

$$2^0) A_k(\theta) = \theta_k^{-1} [A_k(0)] , \quad \forall k \in \mathbb{N}, \quad \forall \theta \in [0, 2\pi] \quad (8)$$

où θ_k désigne l'application $(x_1, \dots, x_k) \quad (\theta x_1, \dots, \theta x_k)$

$$\forall x_1, \dots, x_k \in \mathbb{R}^2 .$$

DÉMONSTRATION.

1°) D'après (5), on a

$$\begin{aligned} \varphi_{\theta}^{P,k}(x_1, \dots, x_k) &= P \left[\bigcup_{i=1}^k \theta^{-1} \right] \longleftarrow , \quad \theta x_i [] \\ &= P \left[\theta^{-1} \bigcup_{i=1}^k \right] \longleftarrow , \quad \theta x_i [] \\ &= (\theta P) \left[\bigcup_{i=1}^k \right] \longleftarrow , \quad \theta x_i [] \end{aligned}$$

$$\forall x_1, \dots, x_k \in \mathbb{R}^2 ; \quad \forall k \in \mathbb{N}. \quad \text{C.Q.F.D.}$$

$$2^0) (x_1, \dots, x_k) \in \theta_k^{-1} [A_k(0)] \iff (\theta x_1, \dots, \theta x_k) \in A_k(0)$$

Cette dernière propriété équivaut à l'existence d'une permutation $(i_1, \dots, i_k)$ de $(1, \dots, k)$ telle que

$$\theta x_{i_1} \leq \dots \leq \theta x_{i_k} \left(\frac{\pi}{2} \right) \text{ donc telle que } x_{i_1} \leq \dots \leq x_{i_k} \left(\theta + \frac{\pi}{2} \right) .$$

THÉORÈME 23 :

Avec les notations précédentes, on a :

$$\text{Prob} [\gamma_n(\theta) = k] = C_n^k \int_{\theta_k^{-1}[A_k(0)]} \varphi_{\theta}^{P,k} [\theta x_1, \dots, \theta x_k]^{n-k} dP^{\otimes k}(x_1, \dots, x_k) \quad (9)$$

$$\forall k = 1, \dots, n .$$

DÉMONSTRATION.

On peut d'abord remarquer que si k points de $\mathbf{R}^2$ sont maximaux modulo θ ils forment une chaîne modulo $(\theta + \frac{\pi}{2})$, ce qui permet d'écrire:

$$\text{Prob} [\gamma_n(\theta) = k] = \text{Prob} [\exists i_1, \dots, i_k \mid X_{i_1}, \dots, X_{i_k} \text{ forment une chaîne modulo } \frac{\pi}{2} + \theta ;$$

$$\forall j \notin \{i_1, \dots, i_k\}, \exists i \in \{i_1, \dots, i_k\} \mid X_j < X_i]$$

On peut toujours supposer que $X_1, \dots, X_k$ sont les k éléments maximaux et écrire :

$$\begin{aligned} \text{Prob} [\gamma_n(\theta) = k] &= C_n^k \text{ Prob} [X_1, \dots, X_k \text{ forment une chaîne modulo} \\ &\quad (\theta + \frac{\pi}{2}) ; \forall j > k, \exists i \leq k \mid X_j < X_i(\theta)] \\ &= C_n^k \int_{A_k(\theta)} \text{Prob} [\forall j > k, \exists i \leq k \mid X_j < x_i(\theta)] dP^{\otimes k}(x_1, \dots, \\ &\quad x_k) \end{aligned}$$

d'où, en tenant compte du fait que les X_i sont indépendants :

$$\text{Prob} [\gamma_n(\theta) = k] = C_n^k \int_{A_k(\theta)} P[\bigcup_{i=1}^k] \leftarrow, x_i[\theta]^{n-k} \cdot dP^{\otimes k}(x_1, \dots, x_k)$$

et le lemme 27 permet d'écrire le résultat sous la forme annoncée.

LEMME 28 :

1°) $\forall x_1, \dots, x_k \in \mathbf{R}^2$, on peut écrire :

$$\varphi_{\theta}^{P,k}(x_1, \dots, x_k) = \sum_{l=1}^n \sum_{i_1, \dots, i_l} (-1)^{l+1} F_{\theta}^P[\inf_{\theta}(x_{i_1}, \dots, x_{i_l}) + 0] \quad (10)$$

2°) Si $x_1 < \dots < x_k$ ($\theta + \frac{\pi}{2}$), on a :

$$\Phi_{\theta}^{P,k}(x_1, \dots, x_k) = \sum_{i=1}^k F_{\theta}^P(x_i) - \sum_{i=1}^{k-1} F_{\theta}^P[\inf_{\theta}(x_i, x_{i+1}) + 0] \quad (11)$$

DÉMONSTRATION.

1°) La première partie du lemme se déduit immédiatement du théorème de Poincaré relatif à la probabilité d'une réunion.

2°) Faisons $\theta = 0$. Il faut montrer que

$$x_1 < \dots < x_k \left(\frac{\pi}{2}\right) \Rightarrow \Phi_0^{P,k}(x_1, \dots, x_k) = \sum_{i=1}^k F_0^P(x_i) - \sum_{i=1}^{k-1} F_0^P(z_i + 0)$$

avec $z_i = \inf(x_i, x_{i+1})$, $\forall i \in \{1, \dots, k-1\}$

La démonstration se fait immédiatement par récurrence en remarquant que, si l'on pose $x_i = (x_{i1}, x_{i2})$, $\forall i = 1, \dots, k$, on a :

$$x_1 < \dots < x_k \left(\frac{\pi}{2}\right) \Rightarrow \begin{cases} x_{11} \leq x_{21} \leq \dots \leq x_{k1} \\ x_{12} \geq x_{22} \geq \dots \geq x_{k2} \end{cases}$$

Pour θ quelconque, on écrit :

$$x_1 < \dots < x_k \left(\theta + \frac{\pi}{2}\right) \iff \theta x_1 < \dots < \theta x_k \left(\frac{\pi}{2}\right)$$

et, compte tenu de (7), on est ramené au cas précédent en remplaçant $\theta x_1, \dots, \theta x_k$ par $x_1, \dots, x_k$, donc :

$$\Phi_{\theta}^{P,k}(x_1, \dots, x_k) = \sum_{i=1}^k F_{\theta}^{\theta P}(\theta x_i) - \sum_{i=1}^{k-1} F_{\theta}^{\theta P}(\inf(\theta x_i, \theta x_{i+1}) + 0)$$

et les relations (2) et (4) permettent de conclure.

COROLLAIRE 21 :

Si F_{θ}^P est continue, on a, $\forall k \in \{1, \dots, n\}$:

$$\text{Prob}[Y_n(\theta) = k] = A_n^k \int_{\theta_k^{-1}(B_k)} \left[\sum_{i=1}^k F_{\theta}^P(x_i) - \sum_{i=1}^{k-1} F_{\theta}^P[z_i(\theta)] \right]^{n-k} \cdot dP^{\otimes k} \\ (x_1, \dots, x_k) \quad (12)$$

avec $A_n^k =$ nombre d'arrangements de n objets k à k .

$$z_i(\theta) = \inf_{\theta} \{x_i, x_{i+1}\}, \quad i = 1, \dots, k-1.$$

$$B_k = \{(x_1, \dots, x_k) \in \mathbb{R}^{2k} \mid x_1 < \dots < x_k \quad (\frac{\pi}{2})\}$$

DÉMONSTRATION.

On peut écrire

$$\theta_k^{-1}(A_k(0)) = \sum_{i_1, \dots, i_k} \theta^{-1} B_{i_1 \dots i_k} + \theta^{-1} A_0$$

où le signe $\sum$ porte sur toutes les permutations de $\{1, \dots, k\}$ et où l'on a posé :

$$B_{i_1, \dots, i_k} = \{(x_1, \dots, x_k) \in A_k(0) \mid x_{i_1} < \dots < x_{i_k} \quad (\frac{\pi}{2})\}$$

et $A_0 = \{(x_1, \dots, x_k) \in A_k(0) \mid i \neq j : x_i = x_j\}$

Or, si F_{θ}^P est continue, F_0^P également donc $\int_{A_0} dF_0^P = 0$.

D'autre part, les événements $B_{i_1, \dots, i_k}$ sont équiprobables, d'où :

$$\int_{\theta_k^{-1}[A_k(0)]} [\varphi_{\theta}^{P,k}]^{n-k} dP^{\otimes k} = k! \int_{\theta_k^{-1}(B_k)} [\varphi_{\theta}^{P,k}]^{n-k} \cdot dP^{\otimes k}$$

Le lemme 28 et le théorème 23 permettent alors de conclure.

REMARQUE :

Si on fait $\theta=0$ (resp. $\theta = \pi$) dans (12), on obtient la loi du nombre d'éléments maximaux (resp. minimaux) pour la relation d'ordre usuelle de $\mathbf{R}^2$.

III - COMPORTEMENT ASYMPTOTIQUE DU NOMBRE D'ÉLÉMENTS MAXIMAUX (RESP. MINIMAUX).

Dans ce paragraphe, nous étudierons le comportement asymptotique du nombre d'éléments extrémaux d'un échantillon de variables aléatoires à valeurs dans $\mathbf{R}^2$ muni de sa relation d'ordre usuelle.

NOTATIONS :

On désignera par Y_n le nombre d'éléments maximaux, par Z_n le nombre d'éléments minimaux et par W_n le nombre d'éléments extrémaux.

1. Espérance Mathématique de Y_n, Z_n, W_n .

THÉORÈME 24 :

Si la loi des variables de l'échantillon est une loi uniforme sur le $[0,1]^2$, on a, quand $n \rightarrow \infty$:

$$1^\circ) \quad E(Y_n) = E(Z_n) \sim \text{Log } n \quad (13)$$

$$2^\circ) \quad E(W_n) \sim 2 \text{Log } n. \quad (14)$$

DÉMONSTRATION.

1°) On peut écrire :

$$\begin{aligned} P_n &= \text{Prob} [X_i \text{ est maximal}] = \text{Prob} [X_j < X_i \text{ ou } X_j \wedge X_i, \forall j \neq i] \\ &= \iint_{[0,1]^2} \text{Prob} [X_j < (x,y) \text{ ou } X_j \wedge (x,y), \forall j \neq i] \, dx \, dy \end{aligned}$$

$$\begin{aligned}
&= \iint_{[0,1]^2} (\text{Prob} [X_j < (x,y) \text{ ou } X_j \wedge (x,y)])^{n-1} dx dy \\
&= \iint_{[0,1]^2} [1 - (1-x)(1-y)]^{n-1} dx dy
\end{aligned}$$

en intégrant, on trouve :

$$P_n = \frac{1}{n} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} \right) \quad (15)$$

Soit alors T_i la variable indicatrice de l'évènement " X_i est maximal" ($i = 1, \dots, n$). Il est clair que

$$Y_n = T_1 + \dots + T_n$$

d'où, d'après (15),

$$E(Y_n) = 1 + \frac{1}{2} + \dots + \frac{1}{n} \quad (16)$$

Par raison de symétrie, on a aussi

$$E(Z_n) = 1 + \frac{1}{2} + \dots + \frac{1}{n} \quad (16 \text{ bis})$$

Donc, quand $n \rightarrow \infty$, on a bien

$$E(Y_n) = E(Z_n) \sim \text{Log } n.$$

2°) Commençons par calculer

$$Q_n = \text{Prob} [X_i \text{ est maximal et minimal}]$$

On a :

$$Q_n = \text{Prob} [X_j \wedge X_i, \forall j \neq i]$$

$$= \iint_{[0,1]^2} [x(1-y) + y(1-x)]^{n-1} dx dy$$

en développant suivant la formule du binôme et en utilisant la fonction B, on trouve :

$$Q_n = \frac{1}{n \cdot n!} \sum_{k=0}^{n-1} k! (n-1-k)! \quad (17)$$

Soit alors V_n le nombre d'éléments à la fois maximaux et minimaux, on aura :

$$E(V_n) = \frac{1}{n!} \sum_{k=0}^{n-1} k! (n-1-k)!$$

ou encore

$$E(V_n) = \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{C_{n-1}^k} \quad (18)$$

En tenant compte du fait que, $\forall k \in \{1, \dots, n-2\}$, on a $C_{n-1}^k \geq n-1$, il vient :

$$E(V_n) \leq \frac{1}{n} \left[2 + \frac{n-2}{n-1} \right] < \frac{3}{n} \quad (19)$$

Or, $W_n = Y_n + Z_n - V_n$, d'où d'après (16), (16 bis) et (18) :

$$E(W_n) = 2 \left(1 + \dots + \frac{1}{n} \right) - \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{C_{n-1}^k} \quad (20)$$

et, d'après (13) et (19), on a bien, quand $n \rightarrow \infty$:

$$E(W_n) \sim 2 \text{Log } n$$

C.Q.F.D.

2. Comportement asymptotique des lois de Y_n et de Z_n .

Posons :

$$P_{n,k} = \text{Prob} [Y_n = k] , \quad \forall n \in \mathbb{N}, \quad \forall k \in \{1, \dots, n\} \quad (21)$$

On a le :

LEMME 29 :

Si la loi des variables de l'échantillon est une loi uniforme sur le carré $[0,1]^2$, on a la relation de récurrence :

$$P_{n,k} = \frac{1}{n} \sum_{h=k-1}^{n-1} P_{h,k-1}, \quad k > 1 \quad (22)$$

DÉMONSTRATION.

Nous allons utiliser la formule de Crofton :

Soit D un domaine $\mathbf{R}^p$ ($p \in \mathbf{N}$) et soit D' un domaine contenant D , la formule de Crofton s'écrit :

$$V \frac{dP}{dV} = n(P_1 - P) \quad (23)$$

où V est le volume de D , P la probabilité d'un certain évènement associé à la figure formée par n points tirés indépendamment et uniformément, dans D et P_1 la probabilité du même évènement quand on tire indépendamment et uniformément $(n-1)$ points dans D et 1 point dans $D'-D$ suivant une loi uniforme dans D' ⁽¹⁾.

Prenons $D = [0, 1]^2$ et $D' = [0, 1 + \varepsilon] \times [0, 1]$, $\varepsilon > 0$. Il est clair que la probabilité de tirer k maximaux en n expériences suivant une loi uniforme est la même dans D et dans D' car l'évènement "il y a k éléments maximaux" ne fait intervenir que l' "ordre" ⁽²⁾.

On a donc $\frac{dP}{dV} = 0$ et (23) s'écrit $P = P_1$. Dans ces conditions, on a successivement :

$$\begin{aligned} P_{n,k} &= \text{Prob} [k \text{ maximaux} \mid (n-1) \text{ points} \in D, 1 \text{ point} \in D'-D] \\ &= \frac{(1+\varepsilon)^n}{n\varepsilon} \text{Prob} [k \text{ maximaux}; (n-1) \text{ points} \in D, 1 \text{ point} \in D'-D] \end{aligned}$$

(1) Pour la démonstration de (23), voir R. DELTHEIL [43] p. 35-37.

(2) Nous verrons plus loin une justification plus précise.

$$= \frac{(1+\varepsilon)^n}{n\varepsilon} \cdot n \iint_{\substack{1 \leq x \leq 1+\varepsilon \\ 0 \leq y \leq 1}} \text{Prob} [k \text{ maximaux; } (n-1) \text{ points } \in D[x,y]] \frac{dx}{1+\varepsilon} \cdot dy$$

Posons $\Delta y = [0,1] \times [y,1]$ et $\Delta' y = [0,1] \times [0,y]$ et désignons par $E_{h,k-1}$ l'évènement "k-1 points sont maximaux parmi les h qui sont tombés sur Δy ".

La loi induite sur Δy par la loi uniforme sur D' étant encore uniforme, on peut écrire :

$$\begin{aligned} \text{Prob}_{\Delta y} [E_{h,k-1}] &= P_{h,k-1} && \text{pour } h \geq k-1 \\ &= 0 && \text{pour } h < k-1 \end{aligned}$$

On en tire alors :

$$\begin{aligned} P_{n,k} &= \frac{(1+\varepsilon)^{n-1}}{\varepsilon} \iint \sum_{\substack{h=k-1 \\ 1 \leq x \leq 1+\varepsilon \\ 0 \leq y \leq 1}}^{n-1} \text{Prob} [h \text{ points } \in \Delta y; (n-1-h) \text{ points } \in \Delta' y; \\ &\quad E_{h,k-1}] \, dx \, dy \\ &= \frac{(1+\varepsilon)^{n-1}}{\varepsilon} \int_1^{1+\varepsilon} dx \sum_{h=k-1}^{n-1} \int_0^1 \frac{C_{n-1}^h (1-y)^h y^{n-1-h}}{(1+\varepsilon)^{n-1}} \cdot P_{h,k-1} \, dy \end{aligned}$$

d'où

$$P_{n,k} = \sum_{h=k-1}^{n-1} C_{n-1}^h P_{h,k-1} B(h+1, n-h)$$

ce qui entraîne

$$P_{n-k} = \frac{1}{n} \sum_{h=k-1}^{n-1} P_{h,k-1}, \quad k > 1$$

C.Q.F.D.

APPLICATIONS :

Si l'on connaît $P_{n,1}$, la relation (22) permet de calculer $P_{n,k}$ pour $k > 1$. Or on a :

$$P_{n,1} = n \text{ Prob } [X_j < \bar{X}_1, \forall j > 1] = n \iint_{[0,1]^2} x^{n-1} y^{n-1} dx dy$$

$$\text{d'où} \quad \boxed{P_{n,1} = \frac{1}{n}} \quad (24)$$

Des calculs faciles montrent alors que :

$$P_{n,n} = \frac{1}{n!}, \quad \forall n \in \mathbb{N} \quad (25)$$

$$P_{n,n-1} = \frac{1}{2(n-2)!}, \quad \forall n \geq 2 \quad (26)$$

$$P_{n,2} = \frac{1}{n} \left(1 + \frac{1}{2} + \dots + \frac{1}{n-1}\right), \quad \forall n \geq 2 \quad (27)$$

Sous les mêmes hypothèses que celles du lemme 29, on a le

THÉORÈME 25 :

Quand $n \rightarrow \infty$, $P_{n,k}$ est équivalent à

$$\frac{(\text{Log } n)^{k-1}}{n \cdot (k-1)!}, \quad \forall k \geq 1 \text{ fixe.}$$

DÉMONSTRATION.

Le résultat est vrai pour $k = 1$ car $P_{n,1} = \frac{1}{n} = \frac{(\text{Log } n)^0}{n \cdot 0!}$; supposons-le vrai pour k et montrons qu'il l'est pour $k+1$.

La relation (22) s'écrira :

$$P_{n,k+1} = \frac{1}{n} \sum_{i=k}^{n-1} P_{i,k}$$

Or. $P_{i,k} = \frac{(\text{Log } i)^{k-1}}{i \cdot (k-1)!} \varepsilon(i,k)$ avec $\lim_{i \rightarrow \infty} \varepsilon(i,k) = 1$.

Donc, $\forall \varepsilon > 0$, il existe $N = N(\varepsilon)$ tel que $i \geq N$ entraîne

$$|1 - \varepsilon(i,k)| < \frac{\varepsilon}{2}$$

On peut toujours choisir $N > e^{k-1}$ et assez grand pour que $N - \frac{(\text{Log } N)^k}{k!}$ soit positif. Ecrivons alors, pour $n > N+2$:

$$P_{n,k+1} = \frac{1}{n} \left[\sum_{i=k}^N P_{i,k} + \sum_{i=N+1}^{n-1} P_{i,k} \right]$$

On en tire

$$P_{n,k+1} < \frac{N}{n} + (1 + \frac{\varepsilon}{2}) \frac{1}{n} \sum_{i=N+1}^{n-1} \frac{(\text{Log } i)^{k-1}}{i \cdot (k-1)!}$$

la fonction $\frac{(\text{Log } x)^{k-1}}{x}$ étant décroissante pour $x > e^{k-1}$, il vient :

$$P_{n,k+1} < \frac{N}{n} + \frac{1 + \frac{\varepsilon}{2}}{n \cdot (k-1)!} \left[\frac{(\text{Log } (n-1))^k}{k} - \frac{(\text{Log } N)^k}{k} \right]$$

d'où

$$P_{n,k+1} < \frac{(1 + \frac{\varepsilon}{2}) (\text{Log } (n-1))^k}{n \cdot k!} + \frac{1}{n} \left[N - \frac{(\text{Log } N)^k}{k!} \right]$$

soit

$$P_{n,k+1} < \frac{(1 + \frac{\varepsilon}{2}) (\text{Log } (n-1))^k}{n \cdot k!} + \frac{\beta_1}{n} \quad (28)$$

où β_1 est, d'après l'hypothèse faite sur N , une quantité positive.

De manière analogue, on peut minorer $P_{n,k+1}$ en écrivant successivement :

$$P_{n,k+1} = \frac{1}{n} \left[\sum_{i=k}^N P_{i,k} + \sum_{i=N+1}^n P_{i,k} \right]$$

$$> \frac{(1-\frac{\epsilon}{2})}{n} \sum_{i=k}^n P_{i,k} + \frac{1}{n} (1-\frac{\epsilon}{2}) \frac{1}{k!} [(\text{Log } (n-1))^k - (\text{Log } (N+1))^k]$$

d'où,

$$P_{n,k+1} > \frac{(\text{Log } (n-1))^k}{n \cdot k!} (1-\frac{\epsilon}{2}) + \frac{\alpha}{n} \quad (29)$$

$$\text{avec } \alpha = -(1-\frac{\epsilon}{2}) \frac{1}{k!} (\text{Log } (N+1))^k = -\beta_2$$

Finalement, (28) et (29) entraînent, $\forall n \geq N$:

$$1 - \frac{\epsilon}{2} - \frac{\beta \cdot k!}{(\text{Log } (n-1))^k} < P_{n,k+1} \cdot \frac{n \cdot k!}{(\text{Log } (n-1))^k} < 1 + \frac{\epsilon}{2} + \frac{\beta \cdot k!}{(\text{Log } (n-1))^k}$$

où l'on a posé $\beta = \sup (\beta_1, \beta_2)$.

On en tire, pour n assez grand,

$$1 - \epsilon < P_{n,k+1} \cdot \frac{n \cdot k!}{(\text{Log } (n-1))^k} < 1 + \epsilon$$

ce qui permet de conclure facilement.

COROLLAIRE 22 :

$$\lim_{n \rightarrow \infty} p. Y_n = +\infty$$

DÉMONSTRATION.

Il suffit de montrer que, $\forall k \in \mathbb{N}$,

$$\text{Prob} [Y_n \leq k] \rightarrow 0$$

$$\text{Or} \quad \text{Prob} [Y_n \leq k] = P_{n,1} + \dots + P_{n,k}$$

donc, d'après le théorème 25 :

$$\text{Prob} [Y_n \leq k] \sim \frac{(\text{Log } n)^{k-1}}{n(k-1)!} \quad (30)$$

quand $n \rightarrow \infty$. C.Q.F.D.

Remarquons que (30) montre que la convergence de Y_n vers l'infini n'est pas presque complètement sûre. On peut se demander si cette convergence est presque sûre.

M. Geffroy a montré qu'il en est bien ainsi :

THÉORÈME 26 : (M. GEFFROY) :

$$\lim_{n \rightarrow \infty} \text{p.s. } Y_n = +\infty.$$

DÉMONSTRATION.

Posons $E_j = \{ Y_j \geq k \}$, $\forall k \in \mathbb{N}$, $\forall j \in \mathbb{N}$.

Il s'agit de montrer que

$$\lim_{n \rightarrow \infty} \text{Prob} \left[\bigcup_{j=n}^{\infty} \overline{E_j} \right] = 0, \quad \forall k \in \mathbb{N}.$$

$$\text{Or} \quad \bigcup_{j=n}^{\infty} \overline{E_j} = \overline{E_n} + \overline{E_{n+1}} \cdot \overline{E_n} + \overline{E_{n+2}} \cdot \overline{E_n} + \overline{E_{n+1}} \cdot \overline{E_n} + \dots$$

$$\text{Donc} \quad \bigcup_{j=n}^{\infty} \overline{E_j} \Rightarrow \overline{E_n} + \overline{E_{n+1}} \cdot \overline{E_n} + \overline{E_{n+2}} \cdot \overline{E_n} + \dots$$

Or, d'après le corollaire 22, on sait que $\lim_{n \rightarrow \infty} \text{Prob}(\overline{E_n}) = 0$. Il suffit donc de montrer que la série de terme général $\text{Prob}[E_n \cdot \overline{E_{n+1}}]$ est convergente.

Soit alors F_{n+1} l'évènement " X_{n+1} est maximal", on a l'application : $E_n \cdot \overline{E_{n+1}} \Rightarrow \overline{E_{n+1}} \cdot F_{n+1}$. D'où :

$$\text{Prob}[E_n \cdot \overline{E_{n+1}}] \leq \text{Prob}[Y_{n+1} < k, X_{n+1} \text{ maximal}]$$

$$\leq \sum_{j=1}^{k-1} \text{Prob}[Y_{n+1} = j] \text{Prob}[X_{n+1} \text{ maximal} | Y_{n+1} = j]$$

$$\leq \sum_{j=1}^{k-1} \text{Prob}[Y_{n+1} = j] \cdot \frac{j}{n+1}$$

$$< \frac{k}{n} \text{Prob}[Y_{n+1} \leq k]$$

Or (30) implique que la série de terme général $\frac{\text{Prob}(Y_{n+1} \leq k)}{n}$ est convergente, ce qui achève la démonstration.

REMARQUE :

Tous les résultats précédents restent valables quand on remplace Y_n par Z_n .

En ce qui concerne W_n , on a le

COROLLAIRE 23 :

$$\lim_{n \rightarrow \infty} \text{p.s. } W_n = +\infty$$

DÉMONSTRATION.

$W_n = Y_n + Z_n - V_n = Y_n + (Z_n - V_n) \geq Y_n$ et l'on conclut à l'aide du théorème 26.

3. Éléments maximaux et éléments saillants,

Résultats complémentaires.

Les résultats qui précèdent montrent qu'il y a une certaine analogie entre les éléments maximaux d'un échantillon à deux dimensions et les éléments saillants d'une suite d'observations ⁽¹⁾. Cette analogie est mise en évidence par le

LEMME 30 :

Dans les conditions du théorème 24, la loi de Y_n est celle du nombre d'éléments saillants d'une suite d'observations.

DÉMONSTRATION.

Soit $f_n(t) = \sum_{k=1}^n e^{itk} P_{n,k}$ la fonction caractéristique de Y_n .

D'après (22), il vient :

$$f_n(t) = \frac{e^{it}}{n} \sum_{k=0}^{n-1} f_k(t)$$

On en tire

$$n f_n(t) - (n-1) f_{n-1}(t) = e^{it} \left[\sum_{k=0}^{n-1} f_k(t) - \sum_{k=0}^{n-2} f_k(t) \right]$$

$$\text{soit } f_n(t) = f_{n-1}(t) \cdot \left[\frac{1}{n} e^{it} + \left(1 - \frac{1}{n}\right) \right]$$

$$\text{d'où } f_n(t) = \prod_{k=1}^n \left[\frac{1}{k} e^{it} + \left(1 - \frac{1}{k}\right) \right]$$

Ce qui signifie que l'on peut écrire

$$\boxed{Y_n = B_1 + \dots + B_n} \quad (31)$$

(1) Pour la définition et les propriétés des éléments saillants d'une suite d'observations, cf RENYI [44].

où B_k ($1 \leq k \leq n$) désigne une variable de Bernoulli d'espérance $\frac{1}{k}$ et où l'on suppose les variables $B_1, \dots, B_n$ indépendantes.

On voit donc que (31) est analogue à la relation (9) de [44] (p. 107). ce qui achève la démonstration.

Interprétation (1) :

Si l'on numérote les points de l'échantillon par abscisses décroissantes, on constate qu'un point est maximal si et seulement si son ordonnée est un élément saillant de la suite des ordonnées: il y a donc le même nombre d'éléments maximaux dans l'échantillon que d'éléments saillants dans la suite des ordonnées des points de l'échantillon.

Remarquons cependant que le processus est différent: le nombre d'éléments saillants d'une suite d'observations est presque sûrement croissant par rapport à la taille de l'échantillon alors qu'il n'en est pas de même pour le nombre d'éléments maximaux.

COROLLAIRE 24 :

$(Y_n - \text{Log } n) / \sqrt{\text{Log } n}$ tend, en loi, vers une variable aléatoire qui suit une loi normale de moyenne nulle et d'écart type 1.

DÉMONSTRATION.

Analogue à celle du théorème 2' de [44].

4. Extensions.

a) Cas où les variables de l'échantillon sont à marges indépendantes.

On a alors le

COROLLAIRE 25 :

Si les variables de l'échantillon sont à marges indépendantes et si les lois marginales sont continues, les résultats précédents restent valables.

{1} Cette interprétation nous a été signalée par M. Geffroy.

DÉMONSTRATION.

Posons $X_i = (X_{i1}^1; X_i^2); i = 1, \dots, n$; et désignons par $F_j (j = 1, 2)$ la loi de X_i .

On sait que $T_i = [F_1 (X_i^1), F_2 (X_i^2)]; i = 1, \dots, n$; suit une loi uniforme sur $[0, 1]^2$: la transformation $(X_1, \dots, X_n) \rightarrow (T_1, \dots, T_n)$ conservant l'ordre, conserve aussi les éléments maximaux, ce qui achève la démonstration.

b) Cas où les marges ne sont pas nécessairement indépendantes.

Si l'on suppose les lois marginales continues on voit, en considérant la transformation définie ci-dessus, que l'on peut toujours se ramener au cas où les variables de l'échantillon sont à valeurs dans $[0, 1]^2$ et sont à lois marginales uniformes sur $[0, 1]$.

D'après un résultat de M. FRECHET [46], on sait alors que la loi H des variables de l'échantillon est telle que :

$$H^0(x, y) \leq H(x, y) \leq H^1(x, y) \quad \forall x, y \in [0, 1] \quad (31)$$

$$\text{avec } H^0(x, y) = \text{Max}[x+y-1, 0] \quad \forall x, y \in [0, 1]$$

$$\text{et } H^1(x, y) = \text{Min}[x, y] \quad \forall x, y \in [0, 1]$$

Pour $H = H^0$, on a $Y_n = n$ p.s., pour $H = H^1$, on a $Y_n = 1$ p.s.

Lorsque H vérifie (31) et qu'elle est absolument continue et à densité continue et positive, on pourrait penser que la loi de Y_n est la même que dans le cas uniforme puisque le nombre d'éléments maximaux ne dépend que de "l'ordre". Le contre-exemple suivant montre qu'il n'en est rien :

Considérons la famille de lois définie par GUMBEL [45] :

$$H_\alpha(x, y) = xy[1 + \alpha(1-x)(1-y)], \quad \forall x, y \in [0, 1] \quad \alpha \in]-1, +1[.$$

Supposons que les variables de l'échantillon ont pour loi H_α et calculons $P_{2,1}(\alpha) = \text{Prob}[Y_2 = 1]$. On aura :

$$P_{2,1}(\alpha) = 2 \iint_{[0,1]^2} xy [1 + \alpha(1-x)(1-y)] [1 + \alpha(1-2x)(1-2y)] dx dy$$

$$\text{d'où l'on tire } P_{2,1}(\alpha) = \frac{1}{2} - \frac{\alpha}{6} \left(1 + \frac{\alpha}{3}\right) \neq P_{2,1} = \frac{1}{2} \text{ pour } \alpha \neq 0.$$

Dans ces conditions, il semble que l'on ne puisse donner que des résultats partiels; on aura, par exemple le :

COROLLAIRE 26 :

Si la loi des variables de l'échantillon s'écrit

$$H(x,y) = xy + \varphi(x,y); \quad \forall x,y \in [0,1] \text{ et si, de plus}$$

$$1^\circ) \varphi(x,y) \leq 0; \quad \forall x,y \in [0,1]$$

$$2^\circ) \varphi \text{ est deux fois dérivable et il existe } A \in]0,1[\text{ tel que}$$

$$\left| \frac{\partial^2 \varphi}{\partial x \partial y}(x,y) \right| \leq A; \quad \forall x,y \in]0,1[$$

alors $E(Y_n)$ est un infiniment grand d'ordre supérieur ou égal à $\text{Log } n$.

DÉMONSTRATION.

On peut écrire $\forall i \in \{1, \dots, n\}$:

$$\text{Prob}[X_i \text{ maximal}] = P_n(H) = \iint_{[0,1]^2} [1 - (1-x)(1-y) - \varphi(x,y)]^{n-1} \cdot$$

$$\left[1 + \frac{\partial^2 \varphi}{\partial x \partial y}(x,y) \right] dx dy$$

d'où $P_n(H) \gg (1-A) P_n$, ce qui permet de conclure en utilisant la relation $E(Y_n) = n P_n(H)$ et le théorème 24.

EXEMPLE :

Les lois H_α vérifient les hypothèses du corollaire 26, $\forall \alpha \in [0,1[$.

Enfin, on peut énoncer le

COROLLAIRE 27 :

$$\text{Si } \lim_{n \rightarrow \infty} P_{n,k} = 0 \quad \text{et} \quad \sum_{n=1}^{\infty} \frac{P_{n,k}}{n} < +\infty ,$$

$$\forall k \in \mathbb{N} \quad \text{alors} \quad \lim_{n \rightarrow \infty} \text{p.s. } Y_n = +\infty .$$

DÉMONSTRATION.

Identique à celle du théorème 26 .

c) Cas d'un échantillon de variables aléatoires à valeurs dans $\mathbb{R}^m$ ($m > 2$).

THÉORÈME 27 :

Si la loi des variables de l'échantillon est la loi uniforme sur $[0,1]^m$, on peut écrire

$$E(Y_n) = E(Z_n) \sim \frac{(\text{Log } n)^{m-1}}{(m-1)!} \quad (32)$$

DÉMONSTRATION.

Posons $P_n(m) = \text{Prob}[X_i \text{ maximal}] = \text{Prob}[X_i \text{ minimal}]$

$\forall i \in \{1, \dots, n\}$. On aura

$$\begin{aligned} P_n(m) &= \int_{[0,1]^m} (1 - x_1 \dots x_m)^{n-1} dx_1 \dots dx_m \\ &= \frac{1}{n} \int_{[0,1]^{m-1}} \left[(1 - x_1 \dots x_m)^n \Big|_{x_m=0}^{x_m=1} \cdot \frac{1}{x_1 \dots x_{m-1}} \right] dx_1 \dots dx_{m-1} \\ &= \frac{1}{n} \int_{[0,1]^{m-1}} \frac{1^n - (1 - x_1 \dots x_{m-1})^n}{x_1 \dots x_{m-1}} dx_1 \dots dx_{m-1} \end{aligned}$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} \int_{[0,1]^{m-1}} (1-x_1 \dots x_{m-1})^k dx_1 \dots dx_{m-1}$$

d'où

$$P_n(m) = \frac{P_1(m-1) + \dots + P_n(m-1)}{n} \quad (33)$$

Or on sait que $P_n(1) = \frac{1}{n}$; on en tire donc

$$P_n(m) \sim \frac{(\text{Log } n)^{m-1}}{n(m-1)!}, \quad \forall m \geq 1 \quad (34)$$

La démonstration s'effectuant de manière tout à fait analogue à celle du théorème 25.

(34) entraîne immédiatement (32), C.Q.F.D.

COROLLAIRE 28 :

Sous les hypothèses du théorème 27, on a

$$\lim_{n \rightarrow \infty} \text{p.s. } Y_n = +\infty.$$

DÉMONSTRATION.

On pose $X_i = (X_i^1, \dots, X_i^m)$, $\forall i \in \{1, \dots, n\}$. Soit alors Y'_n le nombre de points de l'échantillon tels que (X_i^1, X_i^2) soit maximal pour l'échantillon $(X_1^1, X_1^2) \dots (X_n^1, X_n^2)$. On a évidemment $Y'_n < Y_n$, $\forall n \in \mathbb{N}$, or, d'après le théorème 26, $\lim_{n \rightarrow \infty} \text{p.d. } Y'_n = +\infty$.

On a donc aussi $\lim_{n \rightarrow \infty} \text{p.s. } Y_n = +\infty$, C.Q.F.D.

REMARQUE :

Le corollaire 27 reste valable dans $\mathbb{R}^n$ et même dans un espace ordonné quelconque puisque sa démonstration ne nécessite aucune hypothèse

sur la dimension de l'espace dans lequel les variables de l'échantillon prennent leurs valeurs.

IV - THÉORÈME DE M. GEFFROY - APPLICATION A L'ESTIMATION DE LA COURBE MAXIMALE.

1. Théorème de M. Geffroy.

Outre le théorème 26, M. Geffroy a obtenu deux Résultats importants sur le comportement asymptotique du nombre d'éléments maximaux.

Nous donnons seulement les énoncés de ces Résultats; les démonstrations paraîtront prochainement dans une note aux C.R. de M. Geffroy.

Hypothèses et Notations.

On suppose que les variables de l'échantillon sont à valeurs dans $[0,1]^2$ et que leur loi commune est continue et à marges uniformes. On désigne par S le support de la loi des variables de l'échantillon, par S_m l'ensemble des éléments maximaux de S et par M_n l'ensemble des éléments maximaux d'un échantillon de taille n .

THÉORÈME 28 (M. GEFFROY):

$\forall s \in S_m$, $\forall V$ voisinage de s , on a

$$\lim_{n \rightarrow \infty} \text{Prob} \left[\bigcap_{r=n}^{\infty} \{M_r \cap V\} \neq \emptyset \right] = 1.$$

COROLLAIRE 29 (M. GEFFROY):

Si S_m est infini, on a :

$$\lim_{n \rightarrow \infty} \text{p.s. } Y_n = +\infty.$$

DÉFINITION 14 :

Si $s \in S$, on dit que s est non dominant si

$$1^{\circ}) s \in S_m.$$

$$2^{\circ}) \lim_{\substack{h \rightarrow 0 \\ k \rightarrow 0}} \frac{\text{Prob}([s_1-h, s_1] \times [s_2-k, s_2])}{\text{Prob}([s_1-h, s_1] \times [0, s_2-k]) + \text{Prob}([0, s_1-h] \times [s_2-k, s_2])} = 0$$

où l'on a posé $s = (s_1, s_2)$.

THÉORÈME 29 (M. GEFFROY) :

Si S_m est fini, si $s \in S_m$, si V désigne un voisinage de s tel que $V \cap S_m = \{s\}$ et si $Y_n(V)$ désigne le nombre d'éléments maximaux de l'échantillon appartenant à V , alors $\lim_{n \rightarrow \infty} p \cdot Y_n(V) = +\infty$, si et seulement si s est non dominant.

2. Estimation de la courbe maximale.

Nous allons voir comment on peut construire un estimateur convergent de l'ensemble des éléments maximaux de S en utilisant le théorème 28.

Les hypothèses et les notations sont les mêmes que dans le paragraphe précédent. Nous supposons, en outre, que S_m est une courbe continue d'équation $y = g_1(x)$ où g_1 est définie par un intervalle de la forme $[a, 1]$, $0 \leq a \leq 1$, telle que $g_1(a) = 1$.

Posons

$$\begin{cases} g(x) = 1 & \text{pour } x \in [0, a] \\ g(x) = g_1(x) & \text{pour } x \in [a, 1] \end{cases}$$

et soit C_m le graphe de g : nous dirons que C_m est la courbe maximale de S ⁽¹⁾.

(1) Remarquons que g est la fonction frontière de S .

LEMME 31 :

La courbe maximale de S est au-dessus de la droite $y + x - 1 = 0$ et la fonction g est strictement décroissante sur $[a, 1]$.

DÉMONSTRATION.

Soit $F(x, y)$ la f.d.r. de la loi des variables de l'échantillon. On a $(x, y) \in S_m \Rightarrow 1 - x - y + F(x, y) = 0$ d'où $x + y - 1 \geq 0$. D'autre part $(x, y) \in S_m$ et $(x', y') \in S_m$ avec $x' > x$ entraîne bien $y' < y$ puisque (x, y) et (x', y') sont non comparables.

Dans la suite, pour tout $\varepsilon > 0$ et tout entier p , nous poserons :

$$V_\varepsilon = \{ (x, y) \mid x \in [0, 1], \quad g(x) - \varepsilon < y < g(x) + \varepsilon \}$$

et

$$A_{\varepsilon, p, j} = \{ (x, y) \in V_\varepsilon \mid a + j \cdot \frac{1-a}{p} \leq x \leq a + (j+1) \cdot \frac{1-a}{p} \}$$

$$j = 0, 1, \dots, p-1;$$

LEMME 32 :

$\forall \varepsilon > 0, \quad \forall \varepsilon' \in]0, \varepsilon[, \quad p_0 \in \mathbb{N} \quad \text{tel que} \quad \forall p \geq p_0 :$

1°) $\forall M = (x, y), \quad \forall M' = (x', y') \in A_{\varepsilon', p, j}, \quad \forall j \in \{0, \dots, p-1\}$
le segment MM' est inclus dans V_ε .

$$2°) \quad \forall M = (x, y) \in A_{\varepsilon', p, j}, \quad \forall M' = (x', y') \in A_{\varepsilon', p, j+1}.$$

$\forall j \in \{0, \dots, p-2\}$ le segment MM' est inclus dans V_ε .

DÉMONSTRATION.

Pour p assez grand, on aura, d'après la continuité uniforme de g sur $[a, 1]$:

$$\left| g\left(x + 2 \cdot \frac{1-a}{p}\right) - g(x) \right| < \varepsilon - \varepsilon', \quad \forall x \in \left[a, 1 - 2 \cdot \frac{1-a}{p}\right] \quad (35)$$

Soient alors $M_1 = (x_1, y_1)$ et $M_2 = (x_2, y_2) \in A_{\varepsilon', p, j} \cup A_{\varepsilon', p, j+1}$ et tels que $y_2 < y_1$. $\forall M = (x, y)$ et appartenant au segment $M_1 M_2$, on aura $y_2 < y < y_1$, or :

$$y_2 > \inf_{x \in A_{\varepsilon', p, j}} g(x) - \varepsilon' = g(a + (j+1) \cdot \frac{1-a}{p}) - \varepsilon'$$

$$\text{mais } g(a + (j-1) \cdot \frac{1-a}{p}) - g(a + (j+1) \cdot \frac{1-a}{p}) < \varepsilon - \varepsilon'$$

$$\text{d'où } y_2 > g(a + (j-1) \cdot \frac{1-a}{p}) - \varepsilon + \varepsilon' - \varepsilon'$$

$$\text{et } y_2 > \sup_{x \in A_{\varepsilon', p, h} \cup A_{\varepsilon', p, j+1}} g(x) - \varepsilon$$

$$\text{On montrerait de même que } y_1 < \inf_{x \in A_{\varepsilon', p, j} \cup A_{\varepsilon', p, j+1}} g(x) + \varepsilon$$

on a donc bien $M_1 M_2 \subset V_\varepsilon$, C.Q.F.D.

LEMME 33 :

$\forall \varepsilon'' \in]0, \varepsilon'[$ on peut choisir p assez grand pour que $\forall P \in \hat{S} - V_{\varepsilon'}$, $[P, \rightarrow[$ contienne au moins un ensemble $A_{\varepsilon'', p, j}$.

DÉMONSTRATION.

$$\text{Pour } p \text{ assez grand, on a } |g(x + \frac{1-a}{p}) - g(x)| < \frac{\varepsilon' - \varepsilon''}{2},$$

$$\forall x \in [a, 1 - \frac{1-a}{p}]. \text{ Or, si } P = (x_0, y_0) \in \hat{S} - V_{\varepsilon'}, \text{ on a}$$

$y_0 < g(x_0) - \varepsilon'$; alors la droite $y = y_0$ coupe le graphe de $g(x) - \varepsilon''$ ⁽¹⁾ en un tel point que $g(x) = y_0 + \varepsilon''$ donc tel que $g(x) < g(x_0) - \varepsilon' + \varepsilon''$;

(1) Si la droite $y = y_0$ ne coupe pas le graphe de $g(x) - \varepsilon''$, la propriété cherchée est évidente.

ce qui implique $g(x_0) - g(x) > \varepsilon' - \varepsilon''$. Donc, d'après (35) et le fait que g est décroissante, on a :

$$x > x_0 + 2 \cdot \frac{1-a}{p}$$

Ce qui entraîne immédiatement la propriété cherchée.

LEMME 34 :

Soit $M_n(a, p)$ l'ensemble des éléments maximaux de l'échantillon dont l'abscisse appartient à l'intervalle $[a, a - \frac{1-a}{p}]$. Alors, les conditions du lemme 33 étant supposées remplies, on a :

$$\text{Prob} \left[\bigcap_{r=n}^{\infty} M_r(a, p) \subset V_{\varepsilon'} \right] \rightarrow 1 \quad (36)$$

quand $n \rightarrow \infty$.

DÉMONSTRATION.

D'après le théorème 28, on peut écrire :

$$\text{Prob} \left[\bigcap_{r=n}^{\infty} \left[\bigcap_{j=0}^{p-1} \{ M_r \cap A_{\varepsilon'', p, j} \neq \emptyset \} \right] \right] \rightarrow 1$$

mais d'après le lemme 33, l'évènement

$$\left[\bigcap_{j=0}^{p-1} \{ M_r \cap A_{\varepsilon'', p, j} \neq \emptyset \} \right]$$

entraîne qu'un point de l'échantillon appartenant à $\hat{S} - V_{\varepsilon'}$, ne peut être maximal, donc, avec une probabilité 1 que $M_r(a, p) \subset V_{\varepsilon'}$, d'où (36), C.Q.F.D.

Soient maintenant $\mu_1, \dots, \mu_k$ les éléments maximaux d'un échantillon de taille n rangés par abscisses croissantes. Désignons par L_n la ligne brisée $\mu_1 \mu_2 \dots \mu_k$ complétée par des parallèles à $x'ox$.

Soit $\hat{S}_n$ la fonction affine par morceaux dont le graphe est L_n . On a le résultat suivant :

THÉORÈME 30 :

$\hat{g}_n$ converge uniformément presque sûrement vers g sur tout intervalle $[\alpha, \beta]$ tel que $0 < \alpha < \beta < 1$.

DÉMONSTRATION.

Commençons par montrer la convergence sur un intervalle $[\alpha, \beta] \subset]a, 1[$: Soit $\epsilon > 0$ et soit $\epsilon' \in]0, \epsilon]$; choisissons p assez grand pour que les conditions du lemme 32 soient remplies et tel que l'on ait

$$\frac{1-a}{p} < \frac{\alpha-a}{2} \quad \text{et} \quad \frac{1-a}{p} < \frac{1-\beta}{2}.$$

Alors, si $M_n(\alpha, \beta)$ désigne l'ensemble des éléments maximaux dont l'abscisse appartient à $[\alpha, \beta]$,

l'évènement $\bigcap_{r=n}^{\infty} M_r(\alpha, \beta) \subset V_\epsilon$, implique que

$$E_n \cap \{(x, y) \mid \alpha \leq x \leq \beta, \quad y \in \mathbf{R}\} \subset V_\epsilon$$

et le résultat découle immédiatement du lemme 34.

Soit maintenant $h > 0$ donné et soit $\alpha \in]a, 1[$ tel que $g(\alpha) = 1 - h/2$: l'évènement

$$E_n = \bigcap_{r=n}^{\infty} \left\{ \sup_{\alpha \leq x \leq \beta} | \hat{g}_r(x) - g(x) | < \frac{h}{2} \right\}$$

entraîne

$$\bigcap_{r=n}^{\infty} \{ \hat{g}_r(\alpha) > 1 - h \}$$

ce qui, compte tenu du fait que g et $\hat{g}_r$ sont non croissantes, entraîne l'évènement

$$\bigcap_{r=n}^{\infty} \left\{ \sup_{0 \leq x \leq \alpha} | \hat{g}_r(x) - g(x) | < h \right\}$$

Finalement E_n entraîne l'évènement

$$\bigcap_{r=n}^{\infty} \left\{ \sup_{0 \leq x \leq \beta} | \hat{g}_r(x) - g(x) | < h \right\}$$

et comme $\text{Prob}(E_n) \rightarrow 1$, la démonstration est terminée.

CONCLUSION.

Le théorème 30 donne un exemple d'utilisation des éléments maximaux pour estimer une partie de la frontière du support. Nous pensons que, dans des cas plus généraux, il doit être possible d'estimer la frontière du support en utilisant les éléments maximaux associés à une famille de relations d'ordre convenablement choisie.

BIBLIOGRAPHIE

- [33] BHATTACHARYA (P.K.) (1967). Estimation of a Probability density function and its derivatives.
Sankya, Ser. A 29 p. 373-382.
- [36] BHATTACHARYA (P.K.). PARTHASARATHY (K.R.) (1961).
Some limit theorems in regression theory.
Sankya, Ser. A 23 p. 91-102.
- [6] BICKEL (P.J.). LEHMANN (E.L.) (1969). Unbiased estimation in convex families.
Ann. Math. Stat., Vol. 40 n°5 p. 1523-1535.
- [32] BOSQ (D.) (1969). Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales.
Comptes rendus, t. 268 p. 555-557.
- [34] BOSQ (D.) (1969). Estimation non paramétrique de la densité et de ses dérivées.
Comptes rendus, t. 269 p. 1010-1012.
- [37] BOSQ (D.) (1969). Estimation de la densité conditionnelle et de la régression.
Comptes rendus, t. 269 p. 661-664.
- [40] BOSQ (D.) (1968). Sur l'estimation d'une fonction continue.
Comptes rendus, t. 266 p. 837-838.
- [48] BOSQ (D.) (1970). Complément à deux notes sur l'estimation de la densité et de ses dérivées.
Comptes rendus, t. 271 p. 45.
- [42] BOSQ (D.) (1968). Sur les éléments extrémaux d'un échantillon vectoriel.
Comptes rendus, t. 266 p. 570-572.
- [15] BOURBAKI (1955). Espaces vectoriels topologiques (Chapitre IV).
Hermann, Paris.
- [41] BOURBAKI (1952). Intégration (Chapitre IV).
Hermann, Paris.

- [20] CACOULOS (T.) (1966). Estimation of a multivariate density.
Ann. Inst. Stat. Math. 18 p. 179-190.
- [25] CENCOV (N.N.) (1962). Evaluation of an unknown distribution density
from observations.
Soviet Math. 3 p. 1559-1562.
- [2] COURREGÉ (P.) (1965). Théorie de la Mesure.
CDU, Paris.
- [22] CRASWELL (K.J.) (1965). Density estimation in a topological group.
Ann. Math. Stat. 36 p. 1047-1048.
- [43] DELTHEIL (R.) (1926). Probabilités géométriques.
Gauthier-Villars, Paris.
- [9] DUGUE (D.) (1958). Traité de Statistique théorique et appliquée.
Masson, Paris.
- [4] DUNFORD (D.N.), SCHWARTZ (J.T.) (1958). Linear operators (Tome I).
Wiley, New York.
- [13] DVORETZKY (A.), KIEFER (J.), WOLFOWITZ (J.) (1956). Asymptotic
minimax character of the sample distribution function and of the clas-
sical multinomial estimator.
Ann. Math. Stat. 27 p. 642-669.
- [16] FRASER (D.A.S.) (1959). Non parametric Methods in Statistics.
Wiley, New York.
- [46] FRECHET (M.) (1958). Remarques au sujet de la note précédente
([45]).
Comptes rendus, t. 246 p. 2719-2720.
- [39] GEFFROY (J.) (1964). Sur un problème d'estimation géométrique.
Publ. Inst. Stat. Univ. Paris 13 p. 191-210.
- [47] GEFFROY (J.). Sur le nombre d'éléments maximaux d'un échantillon bi-
dimensionnel.
Comptes rendus (à paraître).
- [45] GUMBEL (E.J.) (1958). Distributions à plusieurs variables dont les
marges sont données.
Comptes rendus, t. 246 p. 2717-2719.

- [5] HALMOS (P.R.) (1946). The theory of unbiased estimation.
Ann. Math. Stat. 17 p. 34-43.
- [10] JILOVEC (S.) (1967). On the existence of superconsistent estimates.
4th. conf. Inf. th. Stat. Dec. funct. Band. Proc. Prague 1965 Trans.
Prague, Academia, 24 p. 343-380.
- [28] KRONMAL (R.) et TARTER (M.) (1968). The estimation of probability
densities and cumulatives by Fourier series methods.
J. Amer. Stat. Assoc., 63 p. 925-952.
- [17] LANNUZEL (N.) (1969). Thèse 3ème cycle.
Paris.
- [23] LEADBETTER (M.R.), WATSON (G.S.) (1963). On the estimation of the
probability density I.
Ann. Math. Stat. 34 p. 480-491.
- [49] LE CAM (L.) (1953). On some asymptotic properties of maximum likeli-
hood and related Bayes estimates.
Univ. Calif. Publ. Stat. Vol. 1 p. 277-300.
- [12] LE CAM (L.) (1956). On the asymptotic theory of estimation and testing
hypotheses.
Proc. of 3^d Berkeley symposium p. 129-156
- [11] LE CAM (L.), SCHWARTZ LORRAINE (1960). A necessary and suffi-
cient condition for the existence of consistent estimates.
Ann. Math. Stat. 31 p. 140-150.
- [3] LEHMANN (E.L.) (1951). A general concept of unbiasedness.
Ann. Math. Stat. 22 p. 587-597.
- [31] LOEVE (M.) (1955). Probability Theory.
Van Nostrand, New York.
- [18] NADARAYA (E.A.) (1965). On non-parametric estimates of density
functions and regression curves.
Theor. Prob. Appl. 10 p. 186-190.
- [1] PARZEN (E.) (1958). On asymptotically efficient consistent estimates
of the spectral density of a stationary time series.
J. Roy. Stat. Soc. Ser B 20 p. 303-322.

- [29] PARZEN (E.) (1962). On estimation of a probability density function and mode.
Ann. Math. Stat. 33 p. 1065-1076.
- [24] PICKANDS (J.) III (1969). Efficient estimation of a probability density function.
Ann. Math. Stat. 40 p. 854-864.
- [30] RENYI (A.) (1966). Calcul des Probabilités
Dunod, Paris.
- [44] RENYI (A.) (1962). Théorie des éléments saillants d'une suite d'observations.
Colloquium, Aarhus p. 104-117.
- [38] RENYI (A.), SULANKE (R.) (1963, 1964). Über die convexe hülle von n zufällig gewählten punkten, I et II.
Z. Wahr. 2 p. 75-83 et 3 p. 138-148.
- [7] ROSENBLATT (M.) (1956). Remark on some non parametric estimates of a density function.
Ann. Math. Stat. 27 p. 832-837.
- [35] SANSONE (G.) (1959). Orthogonal functions.
Intersciences, New York.
- [14] SCHUSTER (E.F.) (1969). Estimation of a probability density function and its derivatives.
Ann. Math. Stat. 40 p. 1187-1195.
- [27] SCHWARTZ (S.C.) (1967). Estimation of probability density by an orthogonal series.
Ann. Math. Stat. 38 p. 1261-1265.
- [26] TIAGO DE OLIVEIRA (J.) (1961). Estatística de densidades; Resultados assintóticos.
Rev. Fac. Ci. Univ. Lisboa A 9 1961 p. 165-171.
- [8] TUKEY (J.W.) (1961). Curves as parameters, and touch estimation.
Proc. 4 th. Symp. Math. Stat. Prob.,
Berkeley, p. 681-694.

- [21] VAN RYZIN (J.) (1969). On strong consistency of density estimates.
Ann. Math. Stat. 40 p. 1765-1772.
- [19] WOODROOFE (M.) (1967). On the maximal deviation of the sample density.
Ann. Math. Stat. 38 p. 475-481.

Reçu en Octobre 1970
104. Avenue de Verdun

Esc. 4
94 - IVRY

